Trying 3106016892...Open

Welcome to STN International! Enter x:x
LOGINID:ssspta1644pnh
PASSWORD:
TERMINAL (ENTER 1, 2, 3, OR ?):2

* * * * * * * * * * Welcome to STN International * * * * * * * * * *

NEWS 1		Web Page URLs for STN Seminar Schedule - N. America
NEWS 2	Dec 17	The CA Lexicon available in the CAPLUS and CA files
NEWS 3	Feb 06	Engineering Information Encompass files have new names
NEWS 4	Feb 16	TOXLINE no longer being updated
NEWS 5	Apr 23	Search Derwent WPINDEX by chemical structure
NEWS 6	Apr 23	PRE-1967 REFERENCES NOW SEARCHABLE IN CAPLUS AND CA
NEWS 7	May 07	DGENE Reload
NEWS 8	Jun 20	Published patent applications (A1) are now in USPATFULL
NEWS 9	JUL 13	New SDI alert frequency now available in Derwent's DWPI and DPCI
NEWS 10	Aug 23	In-process records and more frequent updates now in MEDLINE
NEWS 11	Aug 23	PAGE IMAGES FOR 1947-1966 RECORDS IN CAPLUS AND CA
NEWS 12	Aug 23	Adis Newsletters (ADISNEWS) now available on STN
NEWS 13	Sep 17	IMSworld Pharmaceutical Company Directory name change to PHARMASEARCH
NEWS 14	Oct 09	Korean abstracts now included in Derwent World Patents Index
NEWS 15	Oct 09	Number of Derwent World Patents Index updates increased
NEWS 16	Oct 15	Calculated properties now in the REGISTRY/ZREGISTRY File
NEWS 17	Oct 22	Over 1 million reactions added to CASREACT
NEWS 18	Oct 22	DGENE GETSIM has been improved
NEWS 19	Oct 29	AAASD no longer available

NEWS EXPRESS	August 15 CURRENT WINDOWS VERSION IS V6.0c, CURRENT MACINTOSH VERSION IS V6.0 (ENG) AND V6.0J (JP), AND CURRENT DISCOVER FILE IS DATED 07 AUGUST 2001
NEWS HOURS	STN Operating Hours Plus Help Desk Availability
NEWS INTER	General Internet Information
NEWS LOGIN	Welcome Banner and News Items
NEWS PHONE	Direct Dial and Telecommunication Network Access to STN
NEWS WWW	CAS World Wide Web Site (general information)

Enter NEWS followed by the item number or name to see news on that specific topic.

* * * * * * * * * * * * * STN Columbus * * * * * * * * * * * * * * * *

FILE 'HOME' ENTERED AT 14:01:55 ON 05 NOV 2001

=> file medline embase biosis scisearch caplus

COST IN U.S. DOLLARS | SINCE FILE ENTRY | TOTAL SESSION
FULL ESTIMATED COST | 0.15 | 0.15

FILE 'MEDLINE' ENTERED AT 14:02:13 ON 05 NOV 2001

FILE 'EMBASE' ENTERED AT 14:02:13 ON 05 NOV 2001

FILE 'BIOSIS' ENTERED AT 14:02:13 ON 05 NOV 2001

FILE 'SCISEARCH' ENTERED AT 14:02:13 ON 05 NOV 2001

FILE 'CAPLUS' ENTERED AT 14:02:13 ON 05 NOV 2001

=> s grass pollen allergen

L1 1280 GRASS POLLEN ALLERGEN

=> s l1 and hypogenic

L2 0 L1 AND HYPOGENIC

=> s l1 and hypoallergenic

L3 11 L1 AND HYPOALLERGENIC

=> dup remove l3

PROCESSING COMPLETED FOR L3
L4 5 DUP REMOVE L3 (6 DUPLICATES REMOVED)

=> d l4 1-5 cbib abs

L4 ANSWER 1 OF 5 MEDLINE DUPLICATE 1
2001490137 Document Number: 21423541. PubMed ID: 11511525.
Nonanaphylactic synthetic peptides derived from B cell epitopes of the major **grass pollen allergen**, Phl p 1, for allergy vaccination. Focke M; Mahler V; Ball T; Sperr W R; Majlesi Y; Valent P; Kraft D; Valenta R. (Department of Pathophysiology, Vienna General Hospital, AKH, University of Vienna, Waehringer Guertel 18-20, A-1090 Vienna, Austria.) FASEB JOURNAL, (2001 Sep) 15 (11) 2042-4. Journal code: FAS; 8804484. ISSN: 0892-6638. Pub. country: United States. Language: English.
AB Worldwide more than 200 million individuals are allergic to group 1 **grass pollen allergens**. We have used the major timothy **grass pollen allergen** Phl p 1, which cross-reacts with most grass-, corn-, and monocot-derived group 1 allergens to develop a generally applicable strategy for the production of **hypoallergenic** allergy vaccines. On the basis of the experimentally determined B cell epitopes of Phl p 1, we have synthesized five synthetic peptides. These peptides are derived from the major Phl p 1 IgE epitopes and were between 28-32 amino acids long. We demonstrate by

nuclear magnetic resonance that the peptides exhibit no secondary and tertiary structure and accordingly failed to bind IgE antibodies from grass pollen allergic patients. The five peptides, as well as an equimolar mixture thereof, lacked allergenic activity as demonstrated by basophil histamine release and skin test experiments in grass pollen allergic patients. When used as immunogens in mice and rabbits, the peptides induced protective IgG antibodies, which recognized the complete Phl p 1 wild-type allergen and group 1 allergens from other grass species. Moreover, peptide-induced antibodies inhibited the binding of grass pollen allergic patients IgE antibodies to the wild-type allergen. We thus demonstrate that synthetic **hypoallergenic** peptides derived from B cell epitopes of major allergens represent safe vaccine candidates for the treatment of IgE- mediated allergies.

L4 ANSWER 2 OF 5 BIOSIS COPYRIGHT 2001 BIOSIS DUPLICATE 2
2001:178223 Document No.: PREV200100178223. Recombinant **hypoallergenic** fragments of the major timothy **grass pollen allergen**, PHL p 6, for immunotherapy. Vrtala, Susanne (1); Focke, Margit (1); Sperr, Wolfgang; Valent, Peter; Kraft, Dietrich (1); Valenta, Rudolf (1). (1) Institute of Pathophysiology, Vienna Austria. Journal of Allergy and Clinical Immunology, (February, 2001) Vol. 107, No. 2, pp. S257. print. Meeting Info.: 57th Annual Meeting of the American Academy of Allergy, Asthma and Immunology New Orleans, Louisiana, USA March 16-21, 2001 ISSN: 0091-6749. Language: English. Summary Language: English.

L4 ANSWER 3 OF 5 SCISEARCH COPYRIGHT 2001 ISI (R)
2001:645318 The Genuine Article (R) Number: 460PV. Nonanaphylactic synthetic peptides derived from B cell epitopes of the major **grass pollen allergen**, Phl p 1, for allergy vaccination.
Focke M; Mahler V; Ball T; Sperr W R; Majlesi Y; Valent P; Kraft D; Valenta R (Reprint). Univ Vienna, Vienna Gen Hosp, AKH, Dept Pathophysiol, Mol Immunopathol Grp, Waehringer Guertel 18-20, A-1090 Vienna, Austria (Reprint); Univ Vienna, Vienna Gen Hosp, AKH, Dept Pathophysiol, Mol Immunopathol Grp, A-1090 Vienna, Austria; Univ Vienna, Vienna Gen Hosp, AKH, Dept Hematol & Hemostaseol, A-1090 Vienna, Austria; Univ Erlangen Nurnberg, Dept Dermatol, D-8520 Erlangen, Germany. FASEB JOURNAL (JUL 2001) Vol. 15, No. 9, pp. U120-U145. Publisher: FEDERATION AMER SOC EXP BIOL. 9650 ROCKVILLE PIKE, BETHESDA, MD 20814-3998 USA. ISSN: 0892-6638. Pub. country: Austria; Germany. Language: English.
ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS
AB Worldwide more than 200 million individuals are allergic to group 1 **grass pollen allergens**. We have used the major timothy **grass pollen allergen** Phl p 1, which cross-reacts with most grass-, corn-, and monocot-derived group 1 allergens to develop a generally applicable strategy for the production of **hypoallergenic** allergy vaccines. On the basis of the experimentally determined B cell epitopes of Phl p 1, we have synthesized five synthetic peptides. These peptides are derived from the major Phl p 1 IgE epitopes and were between 28-32 amino acids long. We demonstrate by nuclear magnetic resonance that the peptides exhibit no secondary and tertiary structure and accordingly failed to bind IgE antibodies from grass pollen allergic patients. The five peptides, as well as an equimolar mixture thereof, lacked allergenic activity as demonstrated by basophil histamine release and skin test experiments in grass pollen allergic patients. When used as immunogens in mice and rabbits, the peptides

induced protective IgG antibodies, which recognized the complete Phl p 1 wild-type allergen and group 1 allergens from other grass species. Moreover, peptide-induced antibodies inhibited the binding of grass pollen allergic patients IgE antibodies to the wild-type allergen. We thus demonstrate that synthetic **hypoallergenic** peptides derived from B cell epitopes of major allergens represent safe vaccine candidates for the treatment of IgE-mediated allergies.

L4 ANSWER 4 OF 5 SCISEARCH COPYRIGHT 2001 ISI (R)
2001:338197 The Genuine Article (R) Number: 422MU. Reduction in allergenicity of grass pollen by genetic engineering. Bhalla P L (Reprint); Swoboda I; Singh M B. Univ Melbourne, Inst Land & Food Resources, Plant Mol Biol & Biotechnol Lab, Parkville, Vic 3052, Australia (Reprint). INTERNATIONAL ARCHIVES OF ALLERGY AND IMMUNOLOGY (JAN-MAR 2001) Vol. 124, No. 1-3, pp. 51-54. Publisher: KARGER. ALLSCHWILERSTRASSE 10, CH-4009 BASEL, SWITZERLAND. ISSN: 1018-2438. Pub. country: Australia. Language: English.

ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB Background: Hay fever and allergic asthma triggered by **grass pollen allergens** affect similar to 20% of the population in cool temperate climates. Ryegrass is the dominant source of allergens due to its prodigious airborne pollen production. Lol p 5 or group 5 is among the most important and widespread **grass pollen allergen** because it reacts with IgE antibodies of more than 90% of grass pollen-allergic patients, contains most of the grass pollen-specific IgE epitopes and elicits strong biological responses. Significant efforts have been made in developing diagnostic and therapeutic reagents for designing new and more effective immunotherapeutic strategies for treatment of allergic diseases. An alternative approach to this problem could be to reduce the amount of allergen content in the source plant. Methods: High velocity microprojectile bombardment was used to genetically engineer ryegrass. Antisense construct targeted to one of major allergen, Lol p 5, was introduced. The expression of antisense RNA was regulated by a pollen-specific promoter. Pollen was analysed for IgE reactivity. Results: Analysis of proteins with allergen-specific monoclonal and polyclonal antibodies did not detect Lol p 5 in the transgenic pollen. The transgenic pollen showed remarkably reduced allergenicity as reflected by low IgE binding capacity of pollen extract as compared to control pollen. The transgenic ryegrass plants in which Lol p 5 gene expression is perturbed showed normal fertile pollen development. Conclusions: Our studies showed that it is possible to selectively 'switch off' allergen production in pollen of ryegrass demonstrating feasibility of genetic engineering of pla nts for reduced allergenicity. Copyright (C) 2001 S. Karger AG, Basel.

L4 ANSWER 5 OF 5 MEDLINE DUPLICATE 3
1999432292 Document Number: 99432292. PubMed ID: 10500236.
Antisense-mediated silencing of a gene encoding a major ryegrass pollen allergen. Bhalla P L; Swoboda I; Singh M B. (Plant Molecular Biology and Biotechnology Laboratory, Institute of Land and Food Resources, University of Melbourne, Parkville, Victoria 3052, Australia.. p.bhalla@landfood.unimelb.edu.au) . PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, (1999 Sep 28) 96 (20) 11676-80. Journal code: PV3; 7505876. ISSN: 0027-8424. Pub. country: United States. Language: English.

AB Type 1 allergic reactions, such as hay fever and allergic asthma, triggered by **grass pollen allergens** are a

global health problem that affects approximately 20% of the population in cool, temperate climates. Ryegrass is the dominant source of allergens because of its prodigious production of airborne pollen. Lol p 5 is the major allergenic protein of ryegrass pollen, judging from the fact that almost all of the individuals allergic to grass pollen show presence of serum IgE antibodies against this protein. Moreover, nearly two-thirds of the IgE reactivity of ryegrass pollen has been attributed to this protein.

Therefore, it can be expected that down-regulation of Lol p 5 production can significantly reduce the allergic potential of ryegrass pollen. Here, we report down-regulation of Lol p 5 with an antisense construct targeted to the Lol p 5 gene in ryegrass. The expression of antisense RNA was regulated by a pollen-specific promoter. Immunoblot analysis of proteins with allergen-specific antibodies did not detect Lol p 5 in the transgenic pollen. The transgenic pollen showed remarkably reduced allergenicity as reflected by low IgE-binding capacity of pollen extract as compared with that of control pollen. The transgenic ryegrass plants in which Lol p 5 gene expression is perturbed showed normal fertile pollen development, indicating that genetic engineering of **hypoallergenic** grass plants is possible.

=> s l1 and IgE binding

L5 221 L1 AND IGE BINDING

=> s l5 and deletio

L6 0 L5 AND DELETIO

=> s l5 and deletion

L7 8 L5 AND DELETION

=> dup remove l7

PROCESSING COMPLETED FOR L7
L8 2 DUP REMOVE L7 (6 DUPLICATES REMOVED)

=> d l8 1-2 cbib abs

L8 ANSWER 1 OF 2 MEDLINE DUPLICATE 1
2000021846 Document Number: 20021846. PubMed ID: 10553075. Molecular, immunological, and structural characterization of Phl p 6, a major allergen and P-particle-associated protein from Timothy grass (Phleum pratense) pollen. Vrtala S; Fischer S; Grote M; Vangelista L; Pastore A; Sperr W R; Valent P; Reichelt R; Kraft D; Valenta R. (Institute of General and Experimental Pathology, Vienna General Hospital, University of Vienna, Austria.. Susanne.Vrtala@akh-wien.ac.at) . JOURNAL OF IMMUNOLOGY, (1999 Nov 15) 163 (10) 5489-96. Journal code: IFB; 2985117R. ISSN: 0022-1767. Pub. country: United States. Language: English.

AB Due to the wide distribution and heavy pollen production of grasses, approximately 50% of allergic patients are sensitized against **grass pollen allergens**. cDNAs coding for two isoforms and four fragments of a major timothy grass (Phleum pratense) pollen allergen, Phl p 6, were isolated by IgE immunoscreening from a pollen expression cDNA library. Recombinant Phl p 6 (rPhl p 6), an acidic protein of 11.8 kDa, was purified to homogeneity as assessed by mass spectrometry and exhibited almost exclusive alpha-helical secondary structure as determined by circular dichroism spectroscopy. Phl p 6

reacted with serum IgE from 75% of grass pollen-allergic patients (n = 171). **IgE binding** experiments with rPhl p 6 fragments indicated that the N terminus of the allergen is required for IgE recognition. Purified rPhl p 6 elicited dose-dependent basophil histamine release and immediate type skin reactions in patients allergic to grass pollen. A rabbit antiserum raised against purified rPhl p 6 identified it as a pollen-specific protein that, by immunogold electron microscopy, was localized on the polysaccharide-containing wall-precursor bodies (P-particles). The association of Phl p 6 with P-particles may facilitate its intrusion into the deeper airways and thus be responsible for the high prevalence of IgE recognition of Phl p 6. Recombinant native-like Phl p 6 can be used for in vitro as well as in vivo diagnoses of grass pollen allergy, whereas N-terminal **deletion** mutants with reduced **IgE binding** capacity may represent candidates for immunotherapy of grass pollen allergy with a low risk of anaphylactic side effects.

L8 ANSWER 2 OF 2 MEDLINE DUPLICATE 2
1999138945 Document Number: 99138945. PubMed ID: 9973522. "Allergen engineering": variants of the timothy **grass pollen allergen** Phl p 5b with reduced **IgE-binding** capacity but conserved T cell reactivity. Schramm G; Kahlert H; Suck R; Weber B; Stuwe H T; Muller W D; Bufe A; Becker W M; Schlaak M W; Jager L; Cromwell O; Fiebig H. (Biochemische und Molekulare Allergologie, Forschungszentrum Borstel, Germany.. gschramm@fz-borstel.de) . JOURNAL OF IMMUNOLOGY, (1999 Feb 15) 162 (4) 2406-14. Journal code: IFB; 2985117R. ISSN: 0022-1767. Pub. country: United States. Language: English.

AB One problem of conventional allergen-specific immunotherapy is the risk of anaphylactic reactions. A new approach to make immunotherapy safer and more efficient might be the application of engineered allergens with reduced **IgE-binding** capacity but retained T cell reactivity. Using overlapping dodeca-peptides, the dominant T cell epitopes of the timothy **grass pollen allergen** Phl p 5b were identified. By site-directed mutagenesis outside these regions, point and **deletion** mutants were generated. Allergen variants were analyzed for **IgE-binding** capacity with sera of different grass pollen allergic patients by Western blotting, Dot blotting, and EAST inhibition test, and for histamine releasing capacity with peripheral blood basophils from different patients. The **deletion** mutants revealed significantly reduced IgE reactivity and histamine releasing capacity, compared with the wild-type Phl p 5b. Furthermore, in vivo skin prick tests showed that the **deletion** mutants had a significantly lower potency to induce cutaneous reactions than the wild-type Phl p 5b. On the other hand, T cell clones and T cell lines from different allergic patients showed comparable proliferation after stimulation with allergen variants and wild-type Phl p 5b. Considering their reduced anaphylactogenic potential together with their conserved T cell reactivity, the engineered allergens could be important tools for efficient and safe allergen-specific immunotherapy.

=> s l1 and "Phl p6"

L9 0 L1 AND "PHL P6"

=> s l1 and "Phl p6"

L10 0 L1 AND "PHL P6"

=> s (valenta r?/au or vrtala s?/au or stumvoll s?/au or gronlund h?/au or grote m?/au or vangelista l?/au or pastore a?/au or sperr w?/au or valent p?/au or kraft d?/au)

L11 4633 (VALENTA R?/AU OR VRTALA S?/AU OR STUMVOLL S?/AU OR GRONLUND H?/AU OR GROTE M?/AU OR VANGELISTA L?/AU OR PASTORE A?/AU OR SPERR W?/AU OR VALENT P?/AU OR KRAFT D?/AU)

=> s l11 and grass pollen

L12 293 L11 AND GRASS POLLEN

=> dup remove l12

PROCESSING COMPLETED FOR L12
L13 102 DUP REMOVE L12 (191 DUPLICATES REMOVED)

=> s l13 and "p6"

L14 0 L13 AND "P6"

=> s l13 and deletion

L15 2 L13 AND DELETION

=> dup remove l15

PROCESSING COMPLETED FOR L15
L16 2 DUP REMOVE L15 (0 DUPLICATES REMOVED)

=> d l16 1-2 cbib abs

L16 ANSWER 1 OF 2 CAPLUS COPYRIGHT 2001 ACS
2001:319926 Document No. 134:339529 Non-anaphylactic forms of **grass pollen** Phl p 6 allergen and their use. **Valenta, Rudolf**; **Vrtala, Susanne**; Stummvoll, Sabine; Groenlund, Hans; **Grote, Monika**; **Vangelista, Luca**; **Pastore, Annalisa**; **Sperr, Wolfgang R.**; **Valent, Peter**; **Kraft, Dietrich** (Pharmacia Diagnostics Ab, Swed.; et al.). PCT Int. Appl. WO 2001030816 A1 20010503, 43 pp. DESIGNATED STATES: W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF, CG, CH, CI, CM, CY, DE, DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML, MR, NE, NL, PT, SE, SN, TD, TG. (English). CODEN: PIXXD2. APPLICATION: WO 2000-SE2062 20001024. PRIORITY: SE 1999-3950 19991029.

AB The invention relates to a hypoallergenic immunogenic mol. derived from the Phl p 6 allergen, wherein the Phl p 6 mol. has an N-terminal and/or C-terminal **deletion** which makes the mol. at least substantially lack IgE binding capacity. The invention also relates to a hypoallergenic immunogenic combination of mols. derived from the Phl p 6 allergen, comprising: (i) a Phl p 6 mol. having an N-terminal **deletion** which makes the mol. at least substantially lack IgE binding capacity, and (ii) a Phl p 6 mol. having a C-terminal **deletion** which makes the mol. at least substantially lack IgE binding capacity, which two mols. together encompass the complete sequence of Phl p 6. The invention further relates to the use of the hypoallergenic immunogenic mol. or mol. mixt. in hyposensitization and diagnosis.

L16 ANSWER 2 OF 2 MEDLINE
2000021846 Document Number: 20021846. PubMed ID: 10553075. Molecular, immunological, and structural characterization of Phl p 6, a major

allergen and P-particle-associated protein from Timothy grass (Phleum pratense) pollen. **Vrtala S**; Fischer S; **Grote M**; **Vangelista L**; **Pastore A**; **Sperr W R**; **Valent P**; Reichelt R; **Kraft D**; **Valenta R**.
(Institute of General and Experimental Pathology, Vienna General Hospital,
University of Vienna, Austria.. Susanne.Vrtala@akh-wien.ac.at) . JOURNAL OF IMMUNOLOGY, (1999 Nov 15) 163 (10) 5489-96. Journal code: IFB; 2985117R. ISSN: 0022-1767. Pub. country: United States. Language: English.

AB Due to the wide distribution and heavy pollen production of grasses, approximately 50% of allergic patients are sensitized against **grass pollen** allergens. cDNAs coding for two isoforms and four fragments of a major timothy grass (Phleum pratense) pollen allergen, Phl p 6, were isolated by IgE immunoscreening from a pollen expression cDNA library. Recombinant Phl p 6 (rPhl p 6), an acidic protein of 11.8 kDa, was purified to homogeneity as assessed by mass spectrometry and exhibited almost exclusive alpha-helical secondary structure as determined by circular dichroism spectroscopy. Phl p 6 reacted with serum IgE from 75% of **grass pollen**-allergic patients (n = 171). IgE binding experiments with rPhl p 6 fragments indicated that the N terminus of the allergen is required for IgE recognition. Purified rPhl p 6 elicited dose-dependent basophil histamine release and immediate type skin reactions in patients allergic to **grass pollen**. A rabbit antiserum raised against purified rPhl p 6 identified it as a pollen-specific protein that, by immunogold electron microscopy, was localized on the polysaccharide-containing wall-precursor bodies (P-particles). The association of Phl p 6 with P-particles may facilitate its intrusion into the deeper airways and thus be responsible for the high prevalence of IgE recognition of Phl p 6. Recombinant native-like Phl p 6 can be used for in vitro as well as in vivo diagnoses of **grass pollen** allergy, whereas N-terminal **deletion** mutants with reduced IgE binding capacity may represent candidates for immunotherapy of **grass pollen** allergy with a low risk of anaphylactic side effects.

=>

---Logging off of STN---

=>
Executing the logoff script...

=> LOG Y

OM of: US-09-696-169-1 to: Issued_Patents_AA:* out_format : pfs

Date: Nov 5, 2001 7:40 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09696169/runat_05112001_064810_9984/app_query.fasta_1.85
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09696169_@CGN1_1_23 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-696-169-1
Query length: 29
Database: Issued_Patents_AA:*
Database sequences: 212252
Database length: 22503292
Search time (sec): 19.690000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-352-990-27	-	41.00	114.65	13.49	461	!
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-028-366-2	+	41.00	113.54	13.59	527	! S
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-028-366-3	+	41.00	113.54	13.59	527	! S
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-09-136-442-1	+	39.00	116.94	28.77	161	! S
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-924-747-12	-	39.00	114.52	29.26	216	!
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-247-373B-12	-	39.00	114.52	29.26	216	!
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-296-715-12	-	39.00	114.52	29.26	216	!
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-247-373B-6	-	39.00	114.45	29.27	218	!
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-924-747-6	-	39.00	114.41	29.28	219	! S
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-924-747-20	-	39.00	114.41	29.28	219	!
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-247-373B-20	-	39.00	114.41	29.28	219	!
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-296-715-6	-	39.00	114.41	29.28	219	! S
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-296-715-20	-	39.00	114.41	29.28	219	!
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-948-176-26	+	39.00	110.06	30.18	371	!
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-362-670B-26	-	38.00	108.07	45.04	321	!
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-333-576C-26	-	38.00	108.07	45.04	321	!
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-808-324-26	-	38.00	108.07	45.04	321	!
/cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:PCT-US94-14030A-26	-	38.00	108.07	45.04	321	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-196-390-6	+	38.00	100.55	47.46	799	! S
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-811-481-16	-	38.00	98.60	48.11	1012	!
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-854-585-2	+	38.00	96.31	48.88	1337	!
/cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:PCT-US95-05512-2	+	38.00	96.31	48.88	1337	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-188-930-326	-	37.00	104.24	68.09	347	!
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-347-801-4	-	37.00	99.41	70.42	623	! S
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-347-801-18	-	37.00	99.39	70.43	625	!
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-482-728A-7	+	36.00	109.60	96.57	123	!
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-686-878A-14	-	36.00	108.65	97.21	138	!
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-09-136-442-3	+	36.00	108.13	97.56	147	! S
/cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:PCT-US91-00899-14	-	36.00	100.82	102.65	357	
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-09-226-741-3	-	36.00	100.59	102.81	367	! S
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-888-429A-21	+	36.00	100.02	103.22	393	!
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-07-914-281-8	-	36.00	99.78	103.39	405	! S
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-393-246-8	-	36.00	99.78	103.39	405	! S
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-525-058A-8	-	36.00	99.78	103.39	405	!
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-483-151-4	-	36.00	99.78	103.39	405	! S
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-696-731-8	-	36.00	99.78	103.39	405	! S
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-042-531-8	-	36.00	99.78	103.39	405	! S
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-09-226-741-1	-	36.00	98.21	104.53	490	! S
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-906-713-2	-	36.00	96.90	105.48	574	! S
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-545-860D-55	-	36.00	91.59	109.45	1093	!
/cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:PCT-US94-04496-55	-	36.00	91.59	109.45	1	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-190-788A-123	-	35.00	125.60	127.18	12	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-383-474B-128	-	35.00	125.60	127.18	12	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-465-391A-123	-	35.00	125.60	127.18	12	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-464-538B-123	-	35.00	125.60	127.18	12	

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-352-990-27

seq_documentation_block:
; Sequence 27, Application US/09352990
; Patent No. 6255090
; GENERAL INFORMATION:
; APPLICANT: Famodu, Layo O.
; APPLICANT: Orozco, Buddy
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: Plant Aminoacyl-tRNA Synthetase
; FILE REFERENCE: BB-1191
; CURRENT APPLICATION NUMBER: US/09/352,990
; CURRENT FILING DATE: 1999-07-14
; EARLIER APPLICATION NUMBER: 60/092,866
; EARLIER FILING DATE: July 15, 1998
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 27
; LENGTH: 461
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-352-990-27

alignment_scores:
Quality: 41.00 Length: 7
Ratio: 5.857 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 85.714

alignment_block:
US-09-696-169-1/rev x US-09-352-990-27 ..

Align seg 1/1 to: US-09-352-990-27 from: 1 to: 461

```
     27 TCGTGGCCTTCCCCATATGGA 7
        |||||||||||||||:::|||
    193 SerTrpProSerProTrpGly 199
```

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-028-366-2

seq_documentation_block:
; Sequence 2, Application US/09028366
; Patent No. 6150501
; GENERAL INFORMATION:
; APPLICANT: CARLOW, CLOTILDE K.S.
; APPLICANT: HONG, XIQIANG
; APPLICANT: MA, DONG
; TITLE OF INVENTION: NOVEL TYROSINE-CONTAINING
; TITLE OF INVENTION: CYCLOPHILIN AND RELATED METHODS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: New England Biolabs, Inc.
; STREET: 32 Tozer Road
; CITY: Beverly
; STATE: MA
; COUNTRY: US
; ZIP: 01915
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/028,366
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

```
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Williams, Gregory D
;      REGISTRATION NUMBER:  30901
;      REFERENCE/DOCKET NUMBER:  NEB-133
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  978-927-5054
;      TELEFAX:  978-927-1705
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  527 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FRAGMENT TYPE:  internal
US-09-028-366-2


alignment_scores:
            Quality:   41.00             Length:      8
              Ratio:  5.125               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 75.000

alignment_block:
 US-09-696-169-1 x US-09-028-366-2   ..

 Align seg 1/1  to: US-09-028-366-2  from: 1  to: 527

       1 GGGAATTCCATATGGGGAAGGCCA 24
         |||:::|||||||||||:::|||
     343 GlyAspSerIleTrpGlyLysPro 350

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-028-366-3

seq_documentation_block:
; Sequence 3, Application US/09028366
; Patent No. 6150501
;  GENERAL INFORMATION:
;    APPLICANT:  CARLOW, CLOTILDE K.S.
;    APPLICANT:  HONG, XIQIANG
;    APPLICANT:  MA, DONG
;    TITLE OF INVENTION:  NOVEL TYROSINE-CONTAINING
;    TITLE OF INVENTION:  CYCLOPHILIN AND RELATED METHODS
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  New England Biolabs, Inc.
;      STREET:  32 Tozer Road
;      CITY:  Beverly
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  01915
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/028,366
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Williams, Gregory D
;      REGISTRATION NUMBER:  30901
;      REFERENCE/DOCKET NUMBER:  NEB-133
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  978-927-5054
;      TELEFAX:  978-927-1705
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  527 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-028-366-3


alignment_scores:
            Quality:   41.00             Length:      8
              Ratio:  5.125               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 75.000

alignment_block:
 US-09-696-169-1 x US-09-028-366-3   ..

 Align seg 1/1  to: US-09-028-366-3  from: 1  to: 527

       1 GGGAATTCCATATGGGGAAGGCCA 24
         |||:::|||||||||||:::|||
     343 GlyAspSerIleTrpGlyLysPro 350

seq_name: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-09-136-442-1

seq_documentation_block:
; Sequence 1, Application US/09136442
; Patent No. 6030825
; GENERAL INFORMATION:
;  APPLICANT: Hillman, Jennifer L.
;  APPLICANT: Corley, Neil C.
;  APPLICANT: Guegler, Karl J.
;  APPLICANT: Patterson, Chandra
;  TITLE OF INVENTION: CYCLOPHILIN-TYPE PEPTIDYL-PROLYL CIS/TRANS ISOMERASE
;  FILE REFERENCE: PF-0582 US
;  CURRENT APPLICATION NUMBER: US/09/136,442
;  CURRENT FILING DATE: 1998-08-19
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PERL Program
; SEQ ID NO 1
;   LENGTH: 161
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE: -
;   OTHER INFORMATION: 2925455
US-09-136-442-1


alignment_scores:
            Quality:   39.00             Length:      7
              Ratio:  5.571               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 85.714

alignment_block:
 US-09-696-169-1 x US-09-136-442-1   ..

 Align seg 1/1  to: US-09-136-442-1  from: 1  to: 161

       1 GGGAATTCCATATGGGGAAGG 21
         |||||||||||||||||:::
      63 GlyAsnSerIleTrpGlyLys 69

seq_name: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-924-747-12

seq_documentation_block:
; Sequence 12, Application US/08924747
; Patent No. 6063570
;  GENERAL INFORMATION:
;    APPLICANT:  MCGONIGLE, BRIAN
;    APPLICANT:  O'KEEFE, DANIEL
;    TITLE OF INVENTION:  SOYBEAN GLUTATHIONE-S-TRANSFERASE
```

```
;    TITLE OF INVENTION:  ENZYMES
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E.I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.50 INCH
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WORD FOR WINDOWS 95
;      SOFTWARE:  MICROSOFT WORD VERSION 7.0A
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/924,747
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FLOYD, LINDA AXAMETHY
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  CL-1108
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-8112
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  216 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      TISSUE TYPE:  SOYBEAN
;    IMMEDIATE SOURCE:
;      CLONE:  SE6.PK0048.D7
US-08-924-747-12


alignment_scores:
            Quality:   39.00              Length:      7
              Ratio:  5.571                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 71.429

alignment_block:
 US-09-696-169-1/rev x US-08-924-747-12   ..

 Align seg 1/1  to: US-08-924-747-12  from: 1  to: 216

      24 TGGCCTTCCCCATATGGAATT 4
         ||||||||||||:::|||::: 
      11 TrpProSerProPheGlyMet 17

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-247-373B-12

seq_documentation_block:
; Sequence 12, Application US/09247373B
; Patent No. 6168954
; GENERAL INFORMATION:
;  APPLICANT:  MCGONIGLE, BRIAN
;  APPLICANT:  O'KEEFE, DANIEL
;  TITLE OF INVENTION:  SOYBEAN GLUTATHIONE-S-TRANSFERASE ENZYMES
;  FILE REFERENCE:  CL-1108-A
;  CURRENT APPLICATION NUMBER: US/09/247,373B
;  CURRENT FILING DATE:  1999-02-10
;  PRIOR APPLICATION NUMBER:  08/924,747
;  PRIOR FILING DATE:  1997-09-05
;  NUMBER OF SEQ ID NOS:  56
;  SOFTWARE:  Microsoft Office 97
; SEQ ID NO 12
;   LENGTH: 216
;   TYPE: PRT
;   ORGANISM: SOYBEAN
US-09-247-373B-12


alignment_scores:
            Quality:   39.00              Length:      7
              Ratio:  5.571                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 71.429

alignment_block:
 US-09-696-169-1/rev x US-09-247-373B-12   ..

 Align seg 1/1  to: US-09-247-373B-12  from: 1  to: 216

      24 TGGCCTTCCCCATATGGAATT 4
         ||||||||||||:::|||::: 
      11 TrpProSerProPheGlyMet 17

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-296-715-12

seq_documentation_block:
; Sequence 12, Application US/09296715
; Patent No. 6171839
;  GENERAL INFORMATION:
;    APPLICANT:  MCGONIGLE, BRIAN
;    APPLICANT:  O'KEEFE, DANIEL
;    TITLE OF INVENTION:  SOYBEAN GLUTATHIONE-S-TRANSFERASE
;    TITLE OF INVENTION:  ENZYMES
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E.I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.50 INCH
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WORD FOR WINDOWS 95
;      SOFTWARE:  MICROSOFT WORD VERSION 7.0A
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/296,715
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FLOYD, LINDA AXAMETHY
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  CL-1108
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-8112
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  216 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      TISSUE TYPE:  SOYBEAN
;    IMMEDIATE SOURCE:
;      CLONE:  SE6.PK0048.D7
US-09-296-715-12


alignment_scores:
            Quality:   39.00              Length:      7
              Ratio:  5.571                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 71.429

alignment_block:
 US-09-696-169-1/rev x US-09-296-715-12   ..
```

```
 Align seg 1/1  to: US-09-296-715-12  from: 1  to: 216

      24 TGGCCTTCCCCATATGGAATT 4
         ||||||||||||:::|||:::
      11 TrpProSerProPheGlyMet 17

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-247-373B-6

seq_documentation_block:
; Sequence 6, Application US/09247373B
; Patent No. 6168954
; GENERAL INFORMATION:
;  APPLICANT:  MCGONIGLE, BRIAN
;  APPLICANT:  O'KEEFE, DANIEL
;  TITLE OF INVENTION:  SOYBEAN GLUTATHIONE-S-TRANSFERASE ENZYMES
;  FILE REFERENCE:  CL-1108-A
;  CURRENT APPLICATION NUMBER: US/09/247,373B
;  CURRENT FILING DATE:  1999-02-10
;  PRIOR APPLICATION NUMBER:  08/924,747
;  PRIOR FILING DATE:  1997-09-05
;  NUMBER OF SEQ ID NOS:  56
;  SOFTWARE:  Microsoft Office 97
; SEQ ID NO 6
;   LENGTH: 218
;   TYPE: PRT
;   ORGANISM: SOYBEAN
US-09-247-373B-6


alignment_scores:
           Quality:   39.00             Length:      7
             Ratio:  5.571               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 71.429

alignment_block:
 US-09-696-169-1/rev x US-09-247-373B-6   ..

 Align seg 1/1  to: US-09-247-373B-6  from: 1  to: 218

      24 TGGCCTTCCCCATATGGAATT 4
         ||||||||||||:::|||:::
      11 TrpProSerProPheGlyMet 17

seq_name: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-924-747-6

seq_documentation_block:
; Sequence 6, Application US/08924747
; Patent No. 6063570
;  GENERAL INFORMATION:
;    APPLICANT:  MCGONIGLE, BRIAN
;    APPLICANT:  O'KEEFE, DANIEL
;    TITLE OF INVENTION:  SOYBEAN GLUTATHIONE-S-TRANSFERASE
;    TITLE OF INVENTION:  ENZYMES
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E.I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.50 INCH
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WORD FOR WINDOWS 95
;      SOFTWARE:  MICROSOFT WORD VERSION 7.0A
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/924,747
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FLOYD, LINDA AXAMETHY
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  CL-1108
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-8112
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  219 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      TISSUE TYPE:  SOYBEAN
;    IMMEDIATE SOURCE:
;      CLONE:  GSTA
US-08-924-747-6


alignment_scores:
           Quality:   39.00             Length:      7
             Ratio:  5.571               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 71.429

alignment_block:
 US-09-696-169-1/rev x US-08-924-747-6   ..

 Align seg 1/1  to: US-08-924-747-6  from: 1  to: 219

      24 TGGCCTTCCCCATATGGAATT 4
         ||||||||||||:::|||:::
      11 TrpProSerProPheGlyMet 17

seq_name: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-924-747-20

seq_documentation_block:
; Sequence 20, Application US/08924747
; Patent No. 6063570
;  GENERAL INFORMATION:
;    APPLICANT:  MCGONIGLE, BRIAN
;    APPLICANT:  O'KEEFE, DANIEL
;    TITLE OF INVENTION:  SOYBEAN GLUTATHIONE-S-TRANSFERASE
;    TITLE OF INVENTION:  ENZYMES
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E.I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.50 INCH
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WORD FOR WINDOWS 95
;      SOFTWARE:  MICROSOFT WORD VERSION 7.0A
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/924,747
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FLOYD, LINDA AXAMETHY
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  CL-1108
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-8112
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  219 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
```

```
;    ORIGINAL SOURCE:
;      TISSUE TYPE:  SOYBEAN
;    IMMEDIATE SOURCE:
;      CLONE:  SS1.PK0005.E6
US-08-924-747-20


alignment_scores:
            Quality:   39.00              Length:      7
              Ratio:  5.571                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 71.429

alignment_block:
 US-09-696-169-1/rev x US-08-924-747-20    ..

 Align seg 1/1  to: US-08-924-747-20  from: 1  to: 219

      24 TGGCCTTCCCCATATGGAATT 4
         ||||||||||||:::|||:::
      11 TrpProSerProPheGlyMet 17

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-247-373B-20

seq_documentation_block:
; Sequence 20, Application US/09247373B
; Patent No. 6168954
; GENERAL INFORMATION:
;  APPLICANT:  MCGONIGLE, BRIAN
;  APPLICANT:  O'KEEFE, DANIEL
;  TITLE OF INVENTION:  SOYBEAN GLUTATHIONE-S-TRANSFERASE ENZYMES
;  FILE REFERENCE:  CL-1108-A
;  CURRENT APPLICATION NUMBER: US/09/247,373B
;  CURRENT FILING DATE:  1999-02-10
;  PRIOR APPLICATION NUMBER:  08/924,747
;  PRIOR FILING DATE:  1997-09-05
;  NUMBER OF SEQ ID NOS:  56
;  SOFTWARE:  Microsoft Office 97
; SEQ ID NO 20
;   LENGTH: 219
;   TYPE: PRT
;   ORGANISM: SOYBEAN
US-09-247-373B-20


alignment_scores:
            Quality:   39.00              Length:      7
              Ratio:  5.571                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 71.429

alignment_block:
 US-09-696-169-1/rev x US-09-247-373B-20    ..

 Align seg 1/1  to: US-09-247-373B-20  from: 1  to: 219

      24 TGGCCTTCCCCATATGGAATT 4
         ||||||||||||:::|||:::
      11 TrpProSerProPheGlyMet 17

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-296-715-6

seq_documentation_block:
; Sequence 6, Application US/09296715
; Patent No. 6171839
;  GENERAL INFORMATION:
;    APPLICANT:  MCGONIGLE, BRIAN
;    APPLICANT:  O'KEEFE, DANIEL
;    TITLE OF INVENTION:  SOYBEAN GLUTATHIONE-S-TRANSFERASE
;    TITLE OF INVENTION:  ENZYMES
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E.I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.50 INCH
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WORD FOR WINDOWS 95
;      SOFTWARE:  MICROSOFT WORD VERSION 7.0A
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/296,715
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FLOYD, LINDA AXAMETHY
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  CL-1108
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-8112
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  219 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      TISSUE TYPE:  SOYBEAN
;    IMMEDIATE SOURCE:
;      CLONE:  GSTA
US-09-296-715-6


alignment_scores:
            Quality:   39.00              Length:      7
              Ratio:  5.571                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 71.429

alignment_block:
 US-09-696-169-1/rev x US-09-296-715-6    ..

 Align seg 1/1  to: US-09-296-715-6  from: 1  to: 219

      24 TGGCCTTCCCCATATGGAATT 4
         ||||||||||||:::|||:::
      11 TrpProSerProPheGlyMet 17

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-296-715-20

seq_documentation_block:
; Sequence 20, Application US/09296715
; Patent No. 6171839
;  GENERAL INFORMATION:
;    APPLICANT:  MCGONIGLE, BRIAN
;    APPLICANT:  O'KEEFE, DANIEL
;    TITLE OF INVENTION:  SOYBEAN GLUTATHIONE-S-TRANSFERASE
;    TITLE OF INVENTION:  ENZYMES
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E.I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.50 INCH
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WORD FOR WINDOWS 95
;      SOFTWARE:  MICROSOFT WORD VERSION 7.0A
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/296,715
;      FILING DATE:
```

```
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FLOYD, LINDA AXAMETHY
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  CL-1108
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-8112
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  219 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      TISSUE TYPE:  SOYBEAN
;    IMMEDIATE SOURCE:
;      CLONE:  SSl.PK0005.E6
US-09-296-715-20


alignment_scores:
            Quality:   39.00              Length:      7
              Ratio:  5.571                Gaps:       0
 Percent Similarity: 100.000    Percent Identity: 71.429

alignment_block:
 US-09-696-169-1/rev x US-09-296-715-20    ..

 Align seg 1/1  to: US-09-296-715-20  from: 1  to: 219

      24 TGGCCTTCCCCATATGGAATT 4
         |||||||||||:::|||::::
      11 TrpProSerProPheGlyMet 17

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-948-176-26

seq_documentation_block:
; Sequence 26, Application US/08948176
; Patent No. 5945585
;  GENERAL INFORMATION:
;    APPLICANT:  HITZ, WILLIAM D.
;    APPLICANT:  YADAV, NARENDRA S.
;    TITLE OF INVENTION:  ACYL-ACP THIOESTERASES GENES
;    TITLE OF INVENTION:  AND THEIR USE IN ALTERING PLANT
;    TITLE OF INVENTION:  OIL COMPOSITION
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.50 INCH
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WORD FOR WINDOWS 95
;      SOFTWARE:  MICROSOFT WORD VERSION 7.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/948,176
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/631,264
;      FILING DATE:  DECEMBER 20, 1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CHRISTENBURY, LYNNE M.
;      REGISTRATION NUMBER:  30,971
;      REFERENCE/DOCKET NUMBER:  CR-8926-C
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-992-5481
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  371 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-948-176-26


alignment_scores:
            Quality:   39.00              Length:      8
              Ratio:  4.875                Gaps:       0
 Percent Similarity: 100.000    Percent Identity: 62.500

alignment_block:
 US-09-696-169-1 x US-08-948-176-26    ..

 Align seg 1/1  to: US-08-948-176-26  from: 1  to: 371

       4 AATTCCATATGGGGAAGGCCACGA 27
         |||::::::||||||:::||||||
      15 AsnAlaLeuTrpGlyGlnProArg 22

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-362-670B-26

seq_documentation_block:
; Sequence 26, Application US/08362670B
; Patent No. 5658882
;  GENERAL INFORMATION:
;    APPLICANT:  Celeste, Anthony J.
;    APPLICANT:  Wozney, John
;    APPLICANT:  Rosen, Vicki A.
;    APPLICANT:  Wolfman, Neil
;    APPLICANT:  Thomsen, Gerald H.
;    APPLICANT:  Melton, Douglas A.
;    TITLE OF INVENTION:  TENDON-INDUCING COMPOSITIONS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  GENETICS INSTITUTE, INC.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/362,670B
;      FILING DATE:  December 22, 1994
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lazar, Steven R.
;      REGISTRATION NUMBER:  32,618
;      REFERENCE/DOCKET NUMBER:  5202-D
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617 498-8260
;      TELEFAX:  617 876-5851
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  321 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-362-670B-26


alignment_scores:
            Quality:   38.00              Length:      9
              Ratio:  6.333                Gaps:       0
```

Percent Similarity: 66.667 Percent Identity: 66.667

alignment_block:
US-09-696-169-1/rev x US-08-362-670B-26 ..

Align seg 1/1 to: US-08-362-670B-26 from: 1 to: 321

```
  27 TCGTGGCCTTCCCCATATGGAATTCCC 1
     ||||||||   |||   |||   |||
 178 SerTrpProProProSerGlyAlaPro 186
```

```
OM of: US-09-696-169-1 to: SPTREMBL_16:*   out_format : pfs

Date: Nov 5, 2001  7:45 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_n2p.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09696169/runat_05112001_064811_10058/app_query.fasta_1.85
 -DB=SPTREMBL_16 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09696169_@CGN1_1_154 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-696-169-1
Query length: 29
Database: SPTREMBL_16:*
Database sequences: 425026
Database length: 132305027
Search time (sec): 36.120000


score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation  ..
sp_virus:Q85326         +   47.00  148.51   2.93     165  ! Q85326 vaccinia virus. orf g6r.
sp_fungi:Q9Y7D0         -   44.00  131.66  11.58     363  ! Q9y7d0 aspergillus terreus. enc
sp_bacteria:O84756      -   43.00  130.24  17.45     289  ! O84756 chlamydia trachomatis. a
sp_bacteria:Q9PLH2      -   43.00  130.24  17.45     289  ! Q9plh2 chlamydia muridarum. amp
sp_invertebrate:Q22420  +   43.00  128.10  17.88     371  ! Q22420 caenorhabditis elegans
sp_mammal:Q9MZ00        -   43.00  127.31  18.04     407  ! Q9mz00 sus scrofa (pig). beta-3
sp_invertebrate:Q9NJS7  -   43.00  123.68  18.79     622  ! Q9njs7 caenorhabditis elegans
sp_invertebrate:Q9XTK1  -   43.00  122.89  18.96     682  ! Q9xtk1 caenorhabditis elegans
sp_invertebrate:Q9NJS8  -   43.00  122.51  19.04     713  ! Q9njs8 caenorhabditis elegans
sp_invertebrate:Q9U739  -   42.50  128.95  21.56     276  ! Q9u739 podocoryne carnea. hom
sp_virus:O57197         +   42.00  131.66  25.46     165  ! O57197 vaccinia virus (strain a
sp_virus:Q89211         +   42.00  131.66  25.46     165  ! Q89211 variola virus, and varic
sp_bacteria:Q9X167      -   42.00  123.83  27.82     412  ! Q9x167 thermotoga maritima. prc
sp_bacteria:Q9Z859      -   41.00  138.33  35.00      51  ! Q9z859 chlamydia pneumoniae (chl
sp_rodent:Q9Z0W8        -   41.00  130.07  38.44     134  ! Q9z0w8 rattus norvegicus (rat).
sp_plant:Q9ZS19         -   41.00  125.49  40.49     229  ! Q9zs19 alopecurus myosuroides.
sp_plant:Q9ZS20         -   41.00  125.27  40.60     235  ! Q9zs20 alopecurus myosuroides.
sp_plant:Q9SB98         -   41.00  125.27  40.60     235  ! Q9sb98 alopecurus myosuroides.
sp_rodent:Q9ESV7        -   41.00  122.90  41.70     310  ! Q9esv7 rattus norvegicus (rat).
sp_plant:Q9M9F2         -   41.00  119.92  43.13     439  ! Q9m9f2 arabidopsis thaliana (mc
sp_invertebrate:O76824  +   41.00  118.37  43.90     526  ! O76824 brugia malayi. cyclophi
sp_invertebrate:O76964  +   41.00  118.37  43.90     526  ! O76964 onchocerca volvulus. c
sp_invertebrate:O61300  +   41.00  118.36  43.90     527  ! O61300 dirofilaria immitis. c
sp_invertebrate:Q9VQ94  -   41.00  111.38  47.52    1192  ! Q9vq94 drosophila melanogast
sp_invertebrate:Q9NJU7  -   40.50  122.52  51.00     266  ! Q9nju7 helix pomatia (roman s
sp_bacteria:Q53543      -   40.00  130.90  56.47      82  ! Q53543 bacillus halodurans. orf4
sp_bacteria:Q9Z712      -   40.00  120.01  63.90     293  ! Q9z712 chlamydia pneumoniae (ch
sp_bacteria:Q9KD98      -   40.00  115.56  67.21     493  ! Q9kd98 bacillus halodurans. na+
sp_invertebrate:O76826  +   40.00  115.32  67.39     507  ! O76826 caenorhabditis briggsa
sp_bacteria:Q52981      -   40.00  114.67  67.89     547  ! Q52981 rhizobium meliloti (sinc
sp_plant:Q9FJX0         +   40.00  113.95  68.44     595  ! Q9fjx0 arabidopsis thaliana (mc
sp_bacteria:Q9KFY4      +   40.00  111.42  70.44     800  ! Q9kfy4 bacillus halodurans. bhC
sp_virus:Q84515         +   39.00  127.53  87.01      82  ! Q84515 paramecium bursaria chlor
sp_human:Q9H2H8         +   39.00  121.76  92.90     161  ! Q9h2h8 homo sapiens (human). cy
sp_bacteria:Q9KBG1      -   39.00  119.78  95.01     203  ! Q9kbg1 bacillus halodurans. bhl
sp_bacteria:Q9K672      -   39.00  119.57  95.23     208  ! Q9k672 bacillus halodurans. acy
sp_plant:O04874         -   39.00  119.25  95.58     216  ! O04874 glycine max (soybean). g
sp_plant:O49821         -   39.00  119.17  95.67     218  ! O49821 carica papaya (papaya).
sp_plant:O49235         -   39.00  119.13  95.71     219  ! O49235 glycine max 2,4-d induci
sp_plant:Q9FQE8         -   39.00  119.13  95.71     219  ! Q9fqe8 glycine max (soybean). g
sp_plant:P93132         -   39.00  119.09  95.75     220  ! P93132 eucalyptus globulus (
sp_invertebrate:Q9VNH1  -   39.00  116.44  98.68     300  ! Q9vnh1 drosophila melanoga
sp_archea:O52039        -   39.00  114.28  101.12    386  ! O52039 halobacterium sp. (st
sp_plant:O23406         -   39.00  112.82  102.81    458  ! O23406 arabidopsis thaliana
sp_rodent:Q9JKD9        -   39.00  112.69  102.96    465  ! Q9jkd9 mus musculus (mouse).

seq_name: sp_virus:Q85326

seq_documentation_block:
ID   Q85326      PRELIMINARY;      PRT;   165 AA.
AC   Q85326;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   ORF G6R.
OS   Vaccinia virus.
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10245;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220694; PubMed=2024483;
RA   Meis R.J., Condit R.C.;
RT   "Genetic and molecular biological characterization of a vaccinia virus
RT   gene which renders the virus dependent on isatin-beta-
RT   thiosemicarbazone (IBT).";
RL   Virology 182:442-454(1991).
DR   EMBL; J03399; AAB59817.1; -.
SQ   SEQUENCE   165 AA;  18950 MW;  E1032042853384F4 CRC64;


alignment_scores:
           Quality:   47.00              Length:      9
             Ratio:  5.875                Gaps:      0
 Percent Similarity:  88.889   Percent Identity: 88.889

alignment_block:
 US-09-696-169-1 x Q85326   ..

 Align seg 1/1  to: Q85326  from: 1  to: 165

       2 GGAATTCCATATGGGGAAGGCCACGAC 28
         ||||||||||||||   |||||||||
      97 GlyIleProTyrGlyPheGlyHisAsp 105

seq_name: sp_fungi:Q9Y7D0

seq_documentation_block:
ID   Q9Y7D0      PRELIMINARY;      PRT;   363 AA.
AC   Q9Y7D0;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   ENOYL REDUCTASE.
GN   LOVC.
OS   Aspergillus terreus.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX   NCBI_TaxID=33178;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC20542;
RA   Kennedy J., Auclair K., Kendrew S.G., Park C., Vederas J.C.,
RA   Hutchinson C.R.;
RT   "Accessory Proteins Modulate Polyketide Synthase Activity During
RT   Lovastatin Biosynthesis.";
RL   Science 0:0-0(1999).
DR   EMBL; AF141925; AAD34554.1; -.
DR   InterPro; IPR002085; -.
DR   Pfam; PF00107; adh_zinc; 1.
SQ   SEQUENCE   363 AA;  39511 MW;  FDB9524DDB255713 CRC64;
```

```
alignment_scores:
            Quality:   44.00              Length:      9
              Ratio:  5.500                Gaps:      0
 Percent Similarity:  88.889   Percent Identity: 66.667

alignment_block:
 US-09-696-169-1/rev x Q9Y7D0   ..

 Align seg 1/1  to: Q9Y7D0  from: 1  to: 363

      27 TCGTGGCCTTCCCCATATGGAATTCCC 1
         :::|||||||:::|||||||||   |||
     291 ThrTrpProAlaProTyrGlyArgPro 299

seq_name: sp_bacteria:O84756

seq_documentation_block:
ID   O84756      PRELIMINARY;      PRT;   289 AA.
AC   O84756;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   AMP NUCLEOSIDASE.
GN   AMN.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=813;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D/UW-3/CX;
RX   MEDLINE=99000809; PubMed=9784136;
RA   Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA   Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA   Davis R.W.;
RT   "Genome sequence of an obligate intracellular pathogen of humans:
RT   Chlamydia trachomatis.";
RL   Science 282:754-759(1998).
DR   EMBL; AE001347; AAC68346.1; -.
DR   InterPro; IPR000845; -.
DR   Pfam; PF01048; PNP_UDP_1; 1.
SQ   SEQUENCE   289 AA;  32048 MW;  5FB115AE99640CCC CRC64;


alignment_scores:
            Quality:   43.00              Length:      8
              Ratio:  5.375                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 87.500

alignment_block:
 US-09-696-169-1/rev x O84756   ..

 Align seg 1/1  to: O84756  from: 1  to: 289

      26 CGTGGCCTTCCCCATATGGAATTC 3
         :::|||||||||||||||||||||
     264 LysGlyLeuProHisMetGluPhe 271

seq_name: sp_bacteria:Q9PLH2

seq_documentation_block:
ID   Q9PLH2      PRELIMINARY;      PRT;   289 AA.
AC   Q9PLH2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   AMP NUCLEOSIDASE-RELATED PROTEIN.
GN   TC0132.
OS   Chlamydia muridarum.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=83560;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MOPN / NIGG;
RX   MEDLINE=20150255; PubMed=10684935;
RA   Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA   White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA   Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA   Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA   Eisen J., Fraser C.M.;
RT   "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT   pneumoniae AR39.";
RL   Nucleic Acids Res. 28:1397-1406(2000).
DR   EMBL; AE002280; AAF39010.1; -.
DR   TIGR; TC0132; -.
SQ   SEQUENCE   289 AA;  32078 MW;  87AA72E035506745 CRC64;


alignment_scores:
            Quality:   43.00              Length:      8
              Ratio:  5.375                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 87.500

alignment_block:
 US-09-696-169-1/rev x Q9PLH2   ..

 Align seg 1/1  to: Q9PLH2  from: 1  to: 289

      26 CGTGGCCTTCCCCATATGGAATTC 3
         :::|||||||||||||||||||||
     264 LysGlyLeuProHisMetGluPhe 271

seq_name: sp_invertebrate:Q22420

seq_documentation_block:
ID   Q22420      PRELIMINARY;   PRT;   371 AA.
AC   Q22420;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   T12A7.4 PROTEIN.
GN   T12A7.4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lennard N.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z73911; CAA98139.1; -.
DR   InterPro; IPR000379; -.
DR   InterPro; IPR002918; -.
DR   Pfam; PF01674; Lipase_2; 1.
SQ   SEQUENCE   371 AA;  41373 MW;  BE4888C7B37D8EC1 CRC64;


alignment_scores:
            Quality:   43.00              Length:      8
              Ratio:  6.143                Gaps:      0
 Percent Similarity:  87.500   Percent Identity: 75.000
```

```
alignment_block:
 US-09-696-169-1 x Q22420   ..

 Align seg 1/1  to: Q22420  from: 1  to: 371

       1 GGGAATTCCATATGGGGAAGGCCA 24
         ||||||   :::|||||||||||||
     325 GlyAsnMetValTrpGlyArgPro 332

seq_name: sp_mammal:Q9MZ00

seq_documentation_block:
ID   Q9MZ00      PRELIMINARY;      PRT;   407 AA.
AC   Q9MZ00;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   BETA-3-ADRENERGIC RECEPTOR.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Smith T.R., Bidwell C.A., Mills S.E.;
RT   "Sus scrofa beta-3-adrenergic receptor (BAR3) gene.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR   EMBL; AF274007; AAF82301.1; -.
DR   InterPro; IPR000276; -.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
KW   G-protein coupled receptor; Glycoprotein; Receptor; Transmembrane.
SQ   SEQUENCE   407 AA;  43610 MW;  C6598382A9B38DD9 CRC64;


alignment_scores:
            Quality:   43.00             Length:      8
              Ratio:  6.143               Gaps:      0
 Percent Similarity:  87.500   Percent Identity: 75.000

alignment_block:
 US-09-696-169-1/rev x Q9MZ00   ..

 Align seg 1/1  to: Q9MZ00  from: 1  to: 407

      24 TGGCCTTCCCCATATGGAATTCCC 1
         ||||||||||||   |||:::|||
     266 TrpProSerProAlaGlyValPro 273

seq_name: sp_invertebrate:Q9NJS7

seq_documentation_block:
ID   Q9NJS7      PRELIMINARY;      PRT;   622 AA.
AC   Q9NJS7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   G PROTEIN-LINKED ACETYLCHOLINE RECEPTOR GAR-1C.
GN   GAR-1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=20145436; PubMed=10679207;
RA   Park Y.S., Lee Y.S., Cho N.J., Kaang B.K.;
RT   "Alternative splicing of gar-1, a Caenorhabditis elegans G-protein-
RT   linked acetylcholine receptor gene.";
RL   Biochem. Biophys. Res. Commun. 268:354-358(2000).
DR   EMBL; AF117301; AAF26202.1; -.
DR   InterPro; IPR000276; -.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; UNKNOWN_1.
KW   Receptor.
SQ   SEQUENCE   622 AA;  70375 MW;  69A2D47239327B38 CRC64;


alignment_scores:
            Quality:   43.00             Length:      9
              Ratio:  5.375               Gaps:      0
 Percent Similarity:  88.889   Percent Identity: 77.778

alignment_block:
 US-09-696-169-1/rev x Q9NJS7   ..

 Align seg 1/1  to: Q9NJS7  from: 1  to: 622

      27 TCGTGGCCTTCCCCATATGGAATTCCC 1
         ||||||   |||||||||:::||||||
      14 SerTrpAspSerProTyrSerIlePro 22

seq_name: sp_invertebrate:Q9XTK1

seq_documentation_block:
ID   Q9XTK1      PRELIMINARY;      PRT;   682 AA.
AC   Q9XTK1;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   G PROTEIN-LINKED ACETYLCHOLINE RECEPTOR.
GN   GAR-1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99101145; PubMed=9886054;
RA   Lee Y.S., ParK Y.S., Chang D.J., Hwang J.M., Min C.K., Kaang B.K.,
RA   Cho N.J.;
RT   "Cloning and expression of a G protein-linked acetylcholine receptor
RT   from Caenorhabditis elegans.";
RL   J. Neurochem. 72:58-65(1999).
DR   EMBL; AF075245; AAD13747.1; -.
DR   InterPro; IPR000276; -.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; UNKNOWN_1.
KW   Receptor.
SQ   SEQUENCE   682 AA;  76840 MW;  5592614B18EC9517 CRC64;


alignment_scores:
            Quality:   43.00             Length:      9
              Ratio:  5.375               Gaps:      0
 Percent Similarity:  88.889   Percent Identity: 77.778

alignment_block:
 US-09-696-169-1/rev x Q9XTK1   ..

 Align seg 1/1  to: Q9XTK1  from: 1  to: 682

      27 TCGTGGCCTTCCCCATATGGAATTCCC 1
         ||||||   |||||||||:::||||||
      14 SerTrpAspSerProTyrSerIlePro 22

seq_name: sp_invertebrate:Q9NJS8

seq_documentation_block:
```

```
ID   Q9NJS8      PRELIMINARY;      PRT;   713 AA.
AC   Q9NJS8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   G PROTEIN-LINKED ACETYLCHOLINE RECEPTOR GAR-1A.
GN   GAR-1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=20145436; PubMed=10679207;
RA   Park Y.S., Lee Y.S., Cho N.J., Kaang B.K.;
RT   "Alternative splicing of gar-1, a Caenorhabditis elegans G-protein-
RT   linked acetylcholine receptor gene.";
RL   Biochem. Biophys. Res. Commun. 268:354-358(2000).
DR   EMBL; AF117300; AAF26201.1; -.
DR   InterPro; IPR000276; -.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; UNKNOWN_1.
KW   Receptor.
SQ   SEQUENCE   713 AA;  80404 MW;  114EA5065876DEBF CRC64;


alignment_scores:
            Quality:   43.00             Length:      9
              Ratio:  5.375               Gaps:       0
 Percent Similarity:  88.889   Percent Identity: 77.778

alignment_block:
 US-09-696-169-1/rev x Q9NJS8   ..

 Align seg 1/1  to: Q9NJS8  from: 1  to: 713

      27 TCGTGGCCTTCCCCATATGGAATTCCC 1
         ||||||   |||||||||:::||||||
      14 SerTrpAspSerProTyrSerIlePro 22

seq_name: sp_invertebrate:Q9U739

seq_documentation_block:
ID   Q9U739      PRELIMINARY;      PRT;   276 AA.
AC   Q9U739;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   HOMEODOMAIN PROTEIN OTX.
GN   OTX.
OS   Podocoryne carnea.
OC   Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Anthomedusae;
OC   Hydractiniidae; Podocoryne.
OX   NCBI_TaxID=6096;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mueller P., Yanze N., Schmid V., Spring J.;
RT   "The homeobox gene Otx of the jellyfish Podocoryne carnea: role of a
RT   head gene in striated muscle and evolution.";
RL   Dev. Biol. 0:0-0(1999).
DR   EMBL; AF160992; AAF04002.1; -.
DR   HSSP; P06601; 1FJL.
DR   InterPro; IPR001356; -.
DR   Pfam; PF00046; homeobox; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
DR   SMART; SM00389; HOX; 1.
KW   Homeobox; DNA-binding; Nuclear protein.
SQ   SEQUENCE   276 AA;  30662 MW;  B7DD97000F0FE8B4 CRC64;

alignment_scores:
            Quality:   42.50             Length:      9
              Ratio:  5.312               Gaps:       1
 Percent Similarity:  88.889   Percent Identity: 88.889

alignment_block:
 US-09-696-169-1/rev x Q9U739   ..

 Align seg 1/1  to: Q9U739  from: 1  to: 276

      24 TGGCCTTCCCCATAT...GGAATTCCC 1
         |||||||||||||||   |||||||||
      13 TrpProSerProTyrAsnGlyIlePro 21

seq_name: sp_virus:O57197

seq_documentation_block:
ID   O57197      PRELIMINARY;      PRT;   165 AA.
AC   O57197;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   PUTATIVE 18.9K PROTEIN.
GN   MVA076R.
OS   Vaccinia virus (strain Ankara).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=126794;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ANKARA;
RA   Antoine G., Scheiflinger F., Falkner F.G., Dorner F.;
RL   Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U94848; AAB96495.1; -.
SQ   SEQUENCE   165 AA;  18963 MW;  EFE321845455F4FE CRC64;


alignment_scores:
            Quality:   42.00             Length:      9
              Ratio:  5.250               Gaps:       0
 Percent Similarity:  88.889   Percent Identity: 77.778

alignment_block:
 US-09-696-169-1 x O57197   ..

 Align seg 1/1  to: O57197  from: 1  to: 165

       2 GGAATTCCATATGGGGAAGGCCACGAC 28
         |||||||||||||||   ||||||:::
      97 GlyIleProTyrGlyPheGlyHisAsn 105

seq_name: sp_virus:Q89211

seq_documentation_block:
ID   Q89211         PRELIMINARY;   PRT;   165 AA.
AC   Q89211;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   ORF8R.
GN   I7R.
OS   Variola virus, and variola minor virus.
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10255, 53258;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=Variola virus; STRAIN=GARCIA-1966;
RA   Shchelkunov S.N., Sosnovtsev S.V., Totmenin A.V., Resenchuk S.M.,
RA   Blinov V.M., Sandakhchiev L.S.;
RL   Submitted (NOV-1993) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=variola minor virus; STRAIN=GARCIA-1966;
```

```
RA   Shchelkunov S.N., Totmenin A.V., Gutorov V.V., Safronov P.F.,
RA   Massung R.F., Loparev V.N., Knight J.C., Chizhikov V.E., Parsons J.M.,
RA   Esposito J.J., Sosnovtsev S.;
RT   "Analysis of the complete coding sequence of DNA of alastrim variola
RT   minor virus strain Garcia-1966.";
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X76267; CAA53874.1; -.
DR   EMBL; Y16780; CAB54669.1; -.
SQ   SEQUENCE   165 AA;  18962 MW;  EFE329ACF6D7F4FE CRC64;


alignment_scores:
            Quality:   42.00             Length:      9
              Ratio:  5.250               Gaps:      0
 Percent Similarity:  88.889   Percent Identity: 77.778

alignment_block:
 US-09-696-169-1 x Q89211   ..

 Align seg 1/1  to: Q89211  from: 1  to: 165

       2 GGAATTCCATATGGGGAAGGCCACGAC 28
         ||||||||||||||||   ||||||:::
      97 GlyIleProTyrGlyPheGlyHisAsn 105

seq_name: sp_bacteria:Q9X167

seq_documentation_block:
ID   Q9X167      PRELIMINARY;      PRT;   412 AA.
AC   Q9X167;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   PROCESSING PROTEASE, PUTATIVE.
GN   TM1346.
OS   Thermotoga maritima.
OC   Bacteria; Thermotogales; Thermotoga.
OX   NCBI_TaxID=2336;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MSB8 / DSM 3109;
RX   MEDLINE=99287316; PubMed=10360571;
RA   Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA   Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA   McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA   Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA   Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA   Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT   "Evidence for lateral gene transfer between Archaea and Bacteria from
RT   genome sequence of Thermotoga maritima.";
RL   Nature 399:323-329(1999).
DR   EMBL; AE001789; AAD36417.1; -.
DR   TIGR; TM1346; -.
DR   InterPro; IPR001431; -.
DR   Pfam; PF00675; Peptidase_M16; 1.
DR   PROSITE; PS00143; INSULINASE; 1.
KW   Protease.
SQ   SEQUENCE   412 AA;  47413 MW;  960FEFA37AD310E2 CRC64;


alignment_scores:
            Quality:   42.00             Length:      8
              Ratio:  6.000               Gaps:      0
 Percent Similarity:  87.500   Percent Identity: 75.000

alignment_block:
 US-09-696-169-1/rev x Q9X167   ..

 Align seg 1/1  to: Q9X167  from: 1  to: 412

      24 TGGCCTTCCCCATATGGAATTCCC 1
         ||||||:::|||||||||   |||
     145 TrpProGlyProTyrGlyArgPro 152

seq_name: sp_bacteria:Q9Z859

seq_documentation_block:
ID   Q9Z859      PRELIMINARY;      PRT;    51 AA.
AC   Q9Z859;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   HYPOTHETICAL 6.2 KDA PROTEIN.
GN   CPN0492 OR CPJ0492 OR CP0262.
OS   Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX   NCBI_TaxID=83558;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CWL029;
RX   MEDLINE=99206606; PubMed=10192388;
RA   Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA   Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT   "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL   Nat. Genet. 21:385-389(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=J138;
RX   MEDLINE=20330349; PubMed=10871362;
RA   Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA   Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT   "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT   from Japan and CWL029 from USA.";
RL   Nucleic Acids Res. 28:2311-2314(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AR39;
RX   MEDLINE=20150255; PubMed=10684935;
RA   Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA   White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,
RA   Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
RA   Bowman C., Dodson R., Gwinn M., Nelson W., DeBoy R., Kolonay J.,
RA   McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
RT   "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT   pneumoniae AR39.";
RL   Nucleic Acids Res. 28:1397-1406(2000).
DR   EMBL; AE001634; AAD18632.1; -.
DR   EMBL; AP002546; BAA98698.1; -.
DR   EMBL; AE002186; AAF38124.1; -.
DR   TIGR; CP0262; -.
KW   Hypothetical protein.
SQ   SEQUENCE   51 AA;  6154 MW;  D3B06CDA6D46C062 CRC64;


alignment_scores:
            Quality:   41.00             Length:      8
              Ratio:  5.857               Gaps:      0
 Percent Similarity:  87.500   Percent Identity: 75.000

alignment_block:
 US-09-696-169-1/rev x Q9Z859   ..

 Align seg 1/1  to: Q9Z859  from: 1  to: 51

      24 TGGCCTTCCCCATATGGAATTCCC 1
         |||   ||||||||||||:::|||
      35 TrpSerSerProTyrGlyPhePro 42

seq_name: sp_rodent:Q9Z0W8

seq_documentation_block:
ID   Q9Z0W8      PRELIMINARY;      PRT;   134 AA.
AC   Q9Z0W8;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
```

```
DE   CAUDAL-RELATED HOMEOBOX PROTEIN (FRAGMENT).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=ASCENDING COLON;
RA   Dunphy J.L., Taylor R.G., Fuller P.J.;
RT   "Isolation of a partial rat Cdx-2 cDNA.";
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF104031; AAD17915.1; -.
KW   Homeobox; DNA-binding; Nuclear protein.
FT   NON_TER       1      1
FT   NON_TER     134    134
SQ   SEQUENCE   134 AA;  13445 MW;  4014E7985BCDABFF CRC64;


alignment_scores:
            Quality:   41.00             Length:      9
              Ratio:  5.857               Gaps:      0
 Percent Similarity:  77.778   Percent Identity: 66.667

alignment_block:
 US-09-696-169-1/rev x Q9Z0W8   ..

 Align seg 1/1  to: Q9Z0W8  from: 1  to: 134

      27 TCGTGGCCTTCCCCATATGGAATTCCC 1
         |||||||||:::   ||||||   |||
      37 SerTrpProThrAlaTyrGlyAlaPro 45
```

OM of: US-09-696-169-1 to: A_Geneseq_0601:* out_format : pfs

Date: Nov 5, 2001 7:40 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09696169/runat_05112001_064809_9973/app_query.fasta_1.85
-DB=A_Geneseq_0601 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09696169_@CGN1_1_70 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-696-169-1
Query length: 29
Database: A_Geneseq_0601:*
Database sequences: 412676
Database length: 60623988
Search time (sec): 32.770000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY96748	-	44.00	124.90	12.52	363	!
/SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY37396	-	43.00	123.63	18.42	290	!
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAG26982	-	41.00	119.77	40.21	218	!
/SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW53549	-	41.00	116.87	40.88	311	!
/SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB49136	+	41.00	112.57	41.89	527	!
/SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB49137	+	41.00	112.57	41.89	527	!
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAG01708	-	40.00	125.95	56.73	70	! H
/SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY35549	-	40.00	113.75	60.80	312	!
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB38149	-	39.00	126.10	82.82	47	! H
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAG03554	+	39.00	119.78	85.85	102	!
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY78941	+	39.00	116.06	87.69	161	!
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB07831	-	39.00	113.66	88.89	216	!
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB03736	-	39.00	113.66	88.89	216	!
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY79517	-	39.00	113.66	88.89	216	!
/SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB66735	-	39.00	113.66	88.89	216	!
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB07828	-	39.00	113.58	88.93	218	!
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB07835	-	39.00	113.55	88.95	219	!
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB03733	-	39.00	113.55	88.95	219	!
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB03740	-	39.00	113.55	88.95	219	!
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY79514	-	39.00	113.55	88.95	219	!
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY79521	-	39.00	113.55	88.95	219	!
/SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB66732	-	39.00	113.55	88.95	219	!
/SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB66739	-	39.00	113.55	88.95	219	!
/SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY28639	+	39.00	109.24	91.15	371	!
/SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW75021	-	38.00	109.98	132.64	232	!
/SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB76844	-	38.00	108.16	134.02	290	!
/SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:AAR78730	-	38.00	107.33	134.65	321	!
/SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:AAW26591	-	38.00	107.33	134.65	321	!
/SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:AAP50333	+	38.00	105.12	136.36	421	!
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAG31533	-	38.00	104.24	137.04	469	!
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAG31532	-	38.00	103.67	137.49	503	!
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAG31531	-	38.00	103.41	137.69	519	!
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAG50592	-	38.00	103.06	137.96	542	!
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAG50591	-	38.00	102.56	138.35	576	!
/SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY34560	-	38.00	102.49	138.41	581	!
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAG50590	-	38.00	102.34	138.53	592	!
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB37568	+	38.00	102.27	138.58	597	!
/SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY34423	-	38.00	102.21	138.63	601	!
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB37566	+	38.00	99.90	140.46	798	!
/SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW23938	+	38.00	99.89	140.47	799	!
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB37567	+	38.00	99.89	140.47	799	
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB37597	+	38.00	99.89	140.47	799	
/SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:AAW35296	-	38.00	97.96	142.02	1012	
/SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:AAR85203	+	38.00	95.69	143.87	1337	
/SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY59862	-	37.00	110.23	193.54	154	

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY96748

seq_documentation_block:
ID AAY96748 standard; Protein; 363 AA.
XX
AC AAY96748;
XX
DT 09-OCT-2000 (first entry)
XX
DE A. terreus LovC dehydrogenase.
XX
KW Lovastatin; D4B segment; monacolin J; dehydrogenase; LovC; anti-lipemic;
KW HMG-CoA reductase inhibitor; anti-hypercholesterolaemic; anti-fungal.
XX
OS Aspergillus terreus.
XX
PN WO200037629-A2.
XX
PD 29-JUN-2000.
XX
PF 13-DEC-1999; 99WO-US29583.
XX
PR 18-DEC-1998; 98US-0215694.
XX
PA (WISC) WISCONSIN ALUMNI RES FOUND.
XX
PI Hutchinson RC, Kennedy J, Park C;
XX
DR WPI; 2000-442660/38.
DR N-PSDB; AAA51300.
XX
PT Increasing lovastatin or monacolin J production in an organism, for use
PT as antihypercholesterolemic or antifungal agents, comprises
PT transforming the organism with a D4B segment
XX
PS Disclosure; Page 62-63; 116pp; English.
XX
CC The proteins shown in AAY96744-60 are encoded by 17 genes from a
CC cluster in Aspergillus terreus (ATCC 20542), which flank the NPKS
CC (nonaketide polyketide synthase) gene, which is known to be required for
CC lovastatin production. The NPKS gene is contained within the context of
CC the entire gene cluster but is not indicated here (see US5744350). The
CC genes and proteins are named "ORF" or "Lov", where "Lov" signifies genes
CC shown to be essential for lovastatin production. The portion of the gene
CC cluster between ORF1 and the mid-region of LovF is referred to as the
CC "D4B segment". Increasing lovastatin, or monacolin J, production in a
CC lovastatin-producing organism, comprises transforming the organism with
CC a D4B segment, and expressing it (claimed). Lovastatin will also be
CC produced in non-lovastatin producing organisms (e.g. A. nidulans) by
CC transformation with the D4B segment and the entire LovF gene. The
CC methods are used to increase biosynthetic production of lovastatin (with
CC an at least 5-fold increase) which is an anti-hypercholesterolaemic
CC agent, and also has some anti-fungal activity. Lovastatin inhibits the
CC conversion of hydroxymethylglutarylcoenzyme A (HMG-CoA) into mevalonate
CC by HMG-CoA reductase. The methods can also be used to increase production
CC of monacolin J (claimed), which has anti-fungal activity.
XX
SQ Sequence 363 AA;

alignment_scores:
Quality: 44.00 Length: 9
Ratio: 5.500 Gaps: 0
Percent Similarity: 88.889 Percent Identity: 66.667

alignment_block:
US-09-696-169-1/rev x AAY96748 ..

```
 Align seg 1/1  to: AAY96748  from: 1  to: 363

      27 TCGTGGCCTTCCCCATATGGAATTCCC 1
         :::|||||||:::|||||||||   |||
     291 ThrTrpProAlaProTyrGlyArgPro 299

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY37396

seq_documentation_block:
ID   AAY37396 standard; Protein; 290 AA.
XX
AC   AAY37396;
XX
DT   07-OCT-1999  (first entry)
XX
DE   Protein involved in intermediate metabolism of nucleic acids.
XX
KW   Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW   paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;
KW   nongonococcal uretritis; epidymitis; cervicitis; salpingitis;
KW   bartholinitis; pneumopathy; venereal lymphogranulomatosis.
XX
OS   Chlamydia trachomatis.
XX
PN   WO9928475-A2.
XX
PD   10-JUN-1999.
XX
PF   27-NOV-1998;   98WO-IB01939.
XX
PR   04-NOV-1998;   98US-0107077.
PR   28-NOV-1997;   97FR-0015041.
PR   17-DEC-1997;   97FR-0016034.
XX
PA   (GEST ) GENSET.
XX
PI   Griffais R;
XX
DR   WPI; 1999-371125/31.
XX
PT   Genome sequence of Chlamydia trachomatis
XX
PS   Disclosure; Page 1104; 1755pp; English.
XX
CC   AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
CC   of Chlamydia trachomatis (see AAZ01425). The polypeptides can be used as
CC   vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
CC   can also be used to control growth of the microorganism. Chlamydia
CC   trachomatis is responsible for a large number of diseases, e.g. eye
CC   diseases such as conventional trachoma, nonendemic trachoma,
CC   paratrachoma, and inclusion conjunctivitis; genital diseases such as
CC   nongonococcal uretritis, epidymitis, cervicitis, salpingitis,
CC   perihepatitis, bartholinitis; pneumopathy in breast feeding infants;
CC   and venereal lymphogranulomatosis. The polypeptides of the invention
CC   may be of use in treating these diseases.
XX
SQ   Sequence   290 AA;


alignment_scores:
            Quality:   43.00            Length:      8
              Ratio:  5.375              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 87.500

alignment_block:
 US-09-696-169-1/rev x AAY37396   ..

 Align seg 1/1  to: AAY37396  from: 1  to: 290

      26 CGTGGCCTTCCCCATATGGAATTC 3
         :::|||||||||||||||||||||
     265 LysGlyLeuProHisMetGluPhe 272

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAG26982

seq_documentation_block:
ID   AAG26982 standard; Protein; 218 AA.
XX
AC   AAG26982;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 31645.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
```

PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
Quality: 41.00 Length: 7
Ratio: 5.857 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 71.429

alignment_block:
US-09-696-169-1/rev x AAG26982 ..

Align seg 1/1 to: AAG26982 from: 1 to: 218

24 TGGCCTTCCCCATATGGAATT 4
||||||||||||:::|||:::
11 TrpProSerProPheGlyVal 17

```
seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW53549

seq_documentation_block:
ID   AAW53549 standard; Protein; 311 AA.
XX
AC   AAW53549;
XX
DT   23-JUL-1998  (first entry)
XX
DE   Human Cdx2 protein.
XX
KW   Cdx2; Drosophila; caudal protein; human; neoplastic cell; animal model;
KW   intestinal epithelium; carcinoma; colon cancer; detection; diagnosis;
KW   predisposition.
XX
OS   Homo sapiens.
XX
PN   WO9809510-A1.
XX
PD   12-MAR-1998.
XX
PF   01-SEP-1997;   97WO-AU00564.
XX
PR   04-SEP-1996;   96US-0025610.
PR   04-SEP-1996;   96AU-0002108.
PR   04-SEP-1996;   96CA-2184780.
XX
PA   (FLOR-) FLOREY INST EXPERIMENTAL PHYSIOLOGY.
XX
PI   Beck F,  Chawengsaksophak K,  James R;
XX
DR   WPI; 1998-193247/17.
DR   N-PSDB; AAV22213.
XX
PT   Animal model having a Cdx2 Drosophila caudal gene homologue mutation
PT   - useful for developing diagnostic and treatment protocols for colon
PT   cancer
XX
PS   Claim 23; Page 28-29; 49pp; English.
XX
CC   This sequence represents a human homologue of the Drosophila caudal
CC   protein, Cdx2, which is used in a method to genetically alter an animal,
CC   or progeny of the animal, having a predisposition to develop growth of
CC   neoplastic cells in intestinal epithelium. The genetically altered animal
CC   is useful as a model for carcinoma of the colon or a precursor stage of
CC   colon cancer. Cdx2 antibodies are useful for detecting Cdx2 in biological
CC   samples. The presence of a mutation in at least one Cdx2 allele is
CC   indicative of a predisposition to developing familial carcinoma of the
CC   colon or diagnosis of colon cancer. Modulators of Cdx2 are useful for
CC   modulating the expression of Cdx2 in humans. Non-mutated Cdx2 genes can
CC   be used to reduce the likelihood of development of colon cancer or reduce
CC   the spread of colon cancer in a subject.
XX
SQ   Sequence   311 AA;


alignment_scores:
            Quality:   41.00             Length:      9
              Ratio:  5.857               Gaps:      0
 Percent Similarity:  77.778   Percent Identity: 66.667

alignment_block:
 US-09-696-169-1/rev x AAW53549   ..

 Align seg 1/1  to: AAW53549  from: 1  to: 311

      27 TCGTGGCCTTCCCCATATGGAATTCCC 1
         |||||||||:::   ||||||   |||
      64 SerTrpProThrAlaTyrGlyAlaPro 72

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB49136

seq_documentation_block:
ID   AAB49136 standard; Protein; 527 AA.
XX
AC   AAB49136;
XX
DT   08-MAR-2001  (first entry)
XX
DE   D.immitis tyrosine containing cyclophilin.
XX
KW   Cyclophilin; tyrosine; parasite.
XX
OS   Dirofilaria immitis.
XX
PN   US6150501-A.
XX
PD   21-NOV-2000.
XX
PF   24-FEB-1998;   98US-0028366.
XX
PR   24-FEB-1998;   98US-0028366.
XX
PA   (NEWE ) NEW ENGLAND BIOLABS INC.
XX
PI   Carlow CKS,  Ma D,  Hong X;
XX
DR   WPI; 2001-079415/09.
DR   N-PSDB; AAC89355.
XX
PT   Novel cyclophilin, endogenous to e.g. Dirofilaria immitis, useful for
PT   identifying compounds and for treating parasitic infections which are
PT   not susceptible to cyclosporin A, comprises a tyrosine residue in
PT   drug-binding site  -
XX
PS   Claim 2; Fig 1; 28pp; English.
XX
CC   The present invention relates to a cyclophilin including a tyrosine
CC   residue in the drug-binding site, and which is endogenous to the
CC   parasites Onchocerca volvulus, Brugia malayi, Dirofilaria immitis.
CC   The cyclophilin is useful for inhibiting the growth and development
CC   of parasites or for treating parasitic infections which are not
CC   susceptible to cyclosporin A. The purified 'tyrosine-containing'
CC   cyclophilin can be used to produce antibodies, either polyclonal or
CC   monoclonal, useful as probes to detect and/or purify related
CC   cyclophilins in other parasites.
XX
SQ   Sequence   527 AA;


alignment_scores:
            Quality:   41.00             Length:      8
              Ratio:  5.125               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 75.000

alignment_block:
 US-09-696-169-1 x AAB49136   ..

 Align seg 1/1  to: AAB49136  from: 1  to: 527

      1 GGGAATTCCATATGGGGAAGGCCA 24
        |||:::|||||||||||||:::|||
    343 GlyAspSerIleTrpGlyLysPro 350

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB49137

seq_documentation_block:
ID   AAB49137 standard; Protein; 527 AA.
XX
AC   AAB49137;
XX
DT   08-MAR-2001  (first entry)
XX
DE   D.immitis cyclophilin DiCyp-3.
XX
KW   Cyclophilin; tyrosine; parasite.
```

```
XX
OS   Dirofilaria immitis.
XX
PN   US6150501-A.
XX
PD   21-NOV-2000.
XX
PF   24-FEB-1998;   98US-0028366.
XX
PR   24-FEB-1998;   98US-0028366.
XX
PA   (NEWE ) NEW ENGLAND BIOLABS INC.
XX
PI   Carlow CKS,  Ma D,  Hong X;
XX
DR   WPI; 2001-079415/09.
XX
PT   Novel cyclophilin, endogenous to e.g. Dirofilaria immitis, useful for
PT   identifying compounds and for treating parasitic infections which are
PT   not susceptible to cyclosporin A, comprises a tyrosine residue in
PT   drug-binding site -
XX
PS   Disclosure; Fig 2; 28pp; English.
XX
CC   The present invention relates to a cyclophilin including a tyrosine
CC   residue in the drug-binding site, and which is endogenous to the
CC   parasites Onchocerca volvulus, Brugia malayi, Dirofilaria immitis.
CC   The cyclophilin is useful for inhibiting the growth and development
CC   of parasites or for treating parasitic infections which are not
CC   susceptible to cyclosporin A. The purified 'tyrosine-containing'
CC   cyclophilin can be used to produce antibodies, either polyclonal or
CC   monoclonal, useful as probes to detect and/or purify related
CC   cyclophilins in other parasites.
XX
SQ   Sequence   527 AA;


alignment_scores:
           Quality:   41.00            Length:      8
             Ratio:  5.125              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 75.000

alignment_block:
 US-09-696-169-1 x AAB49137   ..

 Align seg 1/1  to: AAB49137  from: 1  to: 527

       1 GGGAATTCCATATGGGGAAGGCCA 24
         |||:::||||||||||||:::|||
     343 GlyAspSerIleTrpGlyLysPro 350

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAG01708

seq_documentation_block:
ID   AAG01708 standard; Protein; 70 AA.
XX
AC   AAG01708;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein, SEQ ID NO: 5789.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   N-PSDB; AAC01714.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 13; SEQ ID 5789; 71pp + CD-ROM; English.
XX
CC   The present sequence is a polypeptide encoded by one of a large number
CC   of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC   were prepared from total human RNAs or polyA+ RNAs derived from 30
CC   different tissues. EST sequences usually correspond mainly to the 3'
CC   untranslated region (UTR) of the mRNA because they are often obtained
CC   from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC   isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC   those cases where longer cDNA sequences have been obtained, the full 5'
CC   UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC   ends and can therefore be used to obtain full length cDNAs and genomic
CC   DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC   chromosome mapping procedures. They are used to obtain upstream
CC   regulatory sequences and to design expression and secretion vectors.
XX
SQ   Sequence   70 AA;


alignment_scores:
           Quality:   40.00            Length:      8
             Ratio:  5.714              Gaps:      0
 Percent Similarity:  87.500   Percent Identity: 75.000

alignment_block:
 US-09-696-169-1/rev x AAG01708   ..

 Align seg 1/1  to: AAG01708  from: 1  to: 70

      24 TGGCCTTCCCCATATGGAATTCCC 1
         ||||||:::|||   |||||||||
      44 TrpProThrProGlyGlyIlePro 51

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY35549

seq_documentation_block:
ID   AAY35549 standard; Protein; 312 AA.
XX
AC   AAY35549;
XX
DT   13-SEP-1999  (first entry)
XX
DE   Protein involved in intermediate metabolism of nucleotides.
XX
KW   Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW   sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW   vaccine; neutralising epitope.
XX
OS   Chlamydia pneumoniae.
XX
PN   WO9927105-A2.
XX
PD   03-JUN-1999.
XX
PF   20-NOV-1998;   98WO-IB01890.
XX
PR   04-NOV-1998;   98US-0107078.
PR   21-NOV-1997;   97FR-0014673.
XX
PA   (GEST ) GENSET.
```

```
XX
PI   Griffais R;
XX
DR   WPI; 1999-357842/30.
XX
PT   Genome sequence of Chlamydia pneumoniae
XX
PS   Page 1296-1297; Disclosure; 1912pp; English.
XX
CC   AAY34584-Y35879 represent the proteins encoded by all the open reading
CC   frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
CC   C. pneumoniae causes respiratory disease such as pneumonia and
CC   bronchitis and is thought to be a contributing factor in heart
CC   disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC   nodosum or pharyngitis. The polypeptides encoded by the open reading
CC   frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
CC   immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC   nucleotides sequences can also be used as immunogenic compositions,
CC   especially where the vector directs the expression of a neutralising
CC   epitope of C. pneumoniae.
XX
SQ   Sequence   312 AA;


alignment_scores:
            Quality:   40.00              Length:      7
              Ratio:  5.714                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-696-169-1/rev x AAY35549   ..

 Align seg 1/1  to: AAY35549  from: 1  to: 312

      26 CGTGGCCTTCCCCATATGGAA 6
         |||||||||||||||||||||
     287 ArgGlyLeuProHisMetGlu 293

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB38149

seq_documentation_block:
ID   AAB38149 standard; Protein; 47 AA.
XX
AC   AAB38149;
XX
DT   30-JAN-2001  (first entry)
XX
DE   Human secreted protein sequence encoded by gene 31 SEQ ID NO:88.
XX
KW   Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW   antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KW   cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW   fungicide; ophthalmological; gene therapy; autoimmune disease; infection;
KW   hyperproliferative disorder; cardiovascular disorder; angiogenesis;
KW   cerebrovascular disorder; nervous system disorder; ocular disorder;
KW   wound healing; skin aging; food additive; preservative.
XX
OS   Homo sapiens.
XX
PN   WO200058468-A2.
XX
PD   05-OCT-2000.
XX
PF   22-MAR-2000; 2000WO-US07526.
XX
PR   26-MAR-1999;   99US-0126600.
PR   22-DEC-1999;   99US-0171550.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM,  Komatsoulis G;
XX
DR   WPI; 2000-611713/58.
DR   N-PSDB; AAC69429.
XX
PT   Nucleic acids encoding human secreted proteins, used to prevent, treat,
PT   ameliorate, or diagnose conditions such as autoimmune disorders, skin
PT   disorders and cancer -
XX
PS   Claim 11; Page 353; 374pp; English.
XX
CC   The polynucleotide sequences given in AAC69399 to AAC69445 encode the
CC   human secreted proteins given in AAB38119 to AAB38165. AAB38166 to
CC   AAB38201 represent human secreted polypeptide sequences and proteins
CC   homologous to them, which are given in the exemplification of the present
CC   invention. Human secreted proteins have activities based on the tissues
CC   and cells the genes are expressed in. Example of activities include:
CC   immunosuppressive; antiarthritic; antirheumatic; antiproliferative;
CC   cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
CC   neuroprotective; antibacterial; virucide; fungicide; and
CC   ophthalmological. The polynucleotides and polypeptides can be are used to
CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC   in diagnosing a pathological condition or susceptibility to a
CC   pathological condition. Disorders which are diagnosed or treated include
CC   autoimmune diseases, hyperproliferative disorders, cardiovascular
CC   disorders, cerebrovascular disorders, angiogenesis, nervous system
CC   disorders, infections caused by bacteria, viruses and fungi and ocular
CC   disorders. The polypeptides can also be used to aid wound healing and
CC   epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC   maintain organs before transplantation, for supporting cell culture of
CC   primary tissues, to regenerate tissues and in chemotaxis. The
CC   polypeptides can also be used as a food additive or preservative to
CC   increase or decrease storage capabilities. AAC69390 to AAC69398 and
CC   AAB38118 represent sequences used in the exemplification of the present
CC   invention.
XX
SQ   Sequence   47 AA;


alignment_scores:
            Quality:   39.00              Length:      9
              Ratio:  4.875                Gaps:      0
 Percent Similarity:  88.889   Percent Identity: 55.556

alignment_block:
 US-09-696-169-1/rev x AAB38149   ..

 Align seg 1/1  to: AAB38149  from: 1  to: 47

      27 TCGTGGCCTTCCCCATATGGAATTCCC 1
         :::||||||   |||:::|||:::|||
      28 AlaTrpProLeuProTrpGlyPhePro 36

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAG03554

seq_documentation_block:
ID   AAG03554 standard; Protein; 102 AA.
XX
AC   AAG03554;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein, SEQ ID NO: 7635.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
```

```
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   N-PSDB; AAC03560.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 13; SEQ ID 7635; 71pp + CD-ROM; English.
XX
CC   The present sequence is a polypeptide encoded by one of a large number
CC   of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC   were prepared from total human RNAs or polyA+ RNAs derived from 30
CC   different tissues. EST sequences usually correspond mainly to the 3'
CC   untranslated region (UTR) of the mRNA because they are often obtained
CC   from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC   isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC   those cases where longer cDNA sequences have been obtained, the full 5'
CC   UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC   ends and can therefore be used to obtain full length cDNAs and genomic
CC   DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC   chromosome mapping procedures. They are used to obtain upstream
CC   regulatory sequences and to design expression and secretion vectors.
XX
SQ   Sequence   102 AA;


alignment_scores:
           Quality:   39.00              Length:      7
             Ratio:  5.571                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 85.714

alignment_block:
 US-09-696-169-1 x AAG03554   ..

 Align seg 1/1  to: AAG03554  from: 1  to: 102

       1 GGGAATTCCATATGGGGAAGG 21
         |||||||||||||||||:::
      63 GlyAsnSerIleTrpGlyLys 69

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY78941

seq_documentation_block:
ID   AAY78941 standard; Protein; 161 AA.
XX
AC   AAY78941;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Cyclophilin-type peptidyl prolyl cis/trans isomerase amino acid sequence.
XX
KW   Cyclophilin-type peptidyl prolyl cis/trans isomerase; CPCI; cancer; AIDS;
KW   leukaemia; reproductive disorder; asthma; diabetes; infertility; anaemia;
KW   polycystic ovary syndrome; uterine fibroid; Good pasture's syndrome;
KW   gout; Grave's disease; multiple sclerosis; lupus; osteoarthritis;
KW   irritable bowel syndrome.
XX
OS   Homo sapiens.
XX
PN   US6030825-A.
XX
PD   29-FEB-2000.
XX
PF   19-AUG-1998;   98US-0136442.
XX
PR   19-AUG-1998;   98US-0136442.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Hillman JL,  Corley NC,  Patterson C,  Guegler KJ;
XX
DR   WPI; 2000-205207/18.
DR   N-PSDB; AAZ95242.
XX
PT   Isolated polynucleotides encoding cyclophilin-type peptidyl-prolyl
PT   cis/trans isomerase, useful for preventing, diagnosing and treating
PT   cancers, autoimmune/inflammatory disorders and reproductive diseases -
XX
PS   Disclosure; Fig 1; 28pp; English.
XX
CC   This sequence represents a human cyclophilin-type peptidyl-prolyl
CC   cis/trans isomerase (CPCI) amino acid sequence. The invention includes
CC   probes for the CPCI nucleotide sequence and vectors expressing the
CC   polynucleotide. CPCI is a member of the peptidyl/prolyl cis/trans
CC   isomerase (PPIase) class of enzymes. Cyclophilin isomerase activity is
CC   essential for correct protein folding and protein trafficking. The CPCI
CC   nucleotide sequence and the protein it encodes may be used in the
CC   diagnosis, prevention and treatment of disorders associated with
CC   inappropriate CPCI expression and activity. For example, they may be used
CC   to treat cancers (e.g. leukaemia, lymphoma, melanoma and cancers of the
CC   breast, liver and prostate), autoimmune/inflammatory disorders
CC   (e.g. AIDS, asthma and diabetes mellitus) and reproductive disorders
CC   (e.g. infertility, polycystic ovary syndrome and uterine fibroids). The
CC   nucleotide sequence may also be used to treat and diagnose allergy,
CC   anaemia, Goodpasture's syndrome, Crohn's disease; gout; Grave's disease,
CC   multiple sclerosis, lupus, irritable bowel syndrome, ulcerative colitis
CC   and osteoarthritis. The CPCI polynucleotide or vectors containing it may
CC   be administered to treat any of the above diseases by rectifying
CC   mutations or deletions in a patient's genome, affecting CPCI metabolism
CC   by expressing inactive proteins or to supplement the patients own
CC   production of CPCI proteins. Conversely, antisense nucleic acid molecules
CC   may be administered to down regulate CPCI protein expression by binding
CC   with the cells own CPCI genes and preventing their expression. Sense and
CC   antisense CPCI nucleotide sequences may also be used as DNA probes in
CC   diagnostic assays (e.g. PCR) to detect and quantitate the presence of
CC   similar nucleic acid sequences in samples, and hence which patients may
CC   be in need of restorative therapy. They may also be used to study the
CC   expression and function of CPCI protein domains and their role in
CC   metabolism.
XX
SQ   Sequence   161 AA;


alignment_scores:
           Quality:   39.00              Length:      7
             Ratio:  5.571                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 85.714

alignment_block:
 US-09-696-169-1 x AAY78941   ..

 Align seg 1/1  to: AAY78941  from: 1  to: 161

       1 GGGAATTCCATATGGGGAAGG 21
         |||||||||||||||||:::
      63 GlyAsnSerIleTrpGlyLys 69

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB07831

seq_documentation_block:
ID   AAB07831 standard; Protein; 216 AA.
XX
AC   AAB07831;
XX
DT   14-NOV-2000  (first entry)
XX
DE   Amino acid sequence of a soybean type III glutathione-S-transferase.
XX
KW   Soybean; glutathione-S-transferase; GST; detoxification;
KW   xenobiotic compound; herbicide-tolerance; transgenic plant;
```

```
KW   herbicide synergist.
XX
OS   Glycine max.
XX
PN   WO200047728-A2.
XX
PD   17-AUG-2000.
XX
PF   10-FEB-2000; 2000WO-US03347.
XX
PR   10-FEB-1999;   99US-0247373.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   McGonigle B,  O'Keefe DP;
XX
DR   WPI; 2000-549144/50.
DR   N-PSDB; AAA59462.
XX
PT   Soybean glutathione-S-transferase polypeptides and polynucleotides used
PT   to produce herbicide tolerant transgenic plants and to screen for
PT   inhibitors or substrates of the enzyme -
XX
PS   Claim 4; Page 61; 84pp; English.
XX
CC   The present sequence represents a soybean glutathione-S-transferase
CC   (GST) enzyme. The enzyme is involved in the detoxification of
CC   xenobiotic compounds in plants and seeds. The GST polynucleotides
CC   and polypeptides are used for the production of herbicide-tolerant
CC   transgenic plants, and for the development of screening assays to
CC   identify GST inhibitors and substrates, which can be used as
CC   herbicide synergists. GST Gene specific probes can be used in gene
CC   identification methods. The recombinant GST enzymes can be used to
CC   produce enzyme specific antibodies which are used to detect the
CC   enzymes in situ in cells or in vitro in cell extracts.
XX
SQ   Sequence   216 AA;


alignment_scores:
            Quality:   39.00              Length:      7
              Ratio:  5.571                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 71.429

alignment_block:
 US-09-696-169-1/rev x AAB07831   ..

 Align seg 1/1  to: AAB07831  from: 1  to: 216

      24 TGGCCTTCCCCATATGGAATT 4
         ||||||||||||:::|||::: 
      11 TrpProSerProPheGlyMet 17

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB03736

seq_documentation_block:
ID   AAB03736 standard; Protein; 216 AA.
XX
AC   AAB03736;
XX
DT   04-OCT-2000  (first entry)
XX
DE   Clone ses8w.pk0028.c6 type III GST protein sequence.
XX
KW   Soybean; glutathione-S-transferase; GST; detoxify; herbicide; stress;
KW   transgenic plant; tolerant; plant breeding.
XX
OS   Glycine max.
XX
PN   US6063570-A.
XX
PD   16-MAY-2000.
XX
PF   05-SEP-1997;   97US-0924747.
XX
PR   05-SEP-1997;   97US-0924747.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   McGonigle B,  O'Keefe DP;
XX
DR   WPI; 2000-375487/32.
DR   N-PSDB; AAA53401.
XX
PT   New Glutathione-S-Transferase enzymes and isolated nucleic acid
PT   fragments encoding them, useful for detoxifying xenobiotic compounds in
PT   plants and seeds, as well as in producing transgenic plants that are
PT   herbicide-resistant  -
XX
PS   Claim 1; Column 41-42; 36pp; English.
XX
CC   This sequence represents a Glutathione-S-Transferase (GST) protein
CC   isolated from a soybean clone. The invention relates to isolated nucleic
CC   acid fragments (see AAA53393-A53406) which encode soybean GST
CC   polypeptides (AAB03731-B03744). GSTs are a family of enzymes which
CC   catalyse the conjugation of glutathione, homoglutathione and other
CC   glutathione-like analogues, to a large range of hydrophobic,
CC   electrophilic compounds. GSTs have been implicated in the detoxification
CC   of certain herbicides. The GST nucleotide sequences are useful in the
CC   construction of herbicide-tolerant transgenic plants, plants that are
CC   tolerant to a wide variety of stresses, or plants in which the GST
CC   enzymes are present at higher or lower levels than they are normally. The
CC   nucleic acid fragments are also useful as probes for genetically and
CC   physically mapping the genes that they are part of, and as markers for
CC   traits linked to expression of the enzymes. This will be useful in plant
CC   breeding in order to develop lines with desired phenotypes or in the
CC   identification of mutants. The soybean GST enzymes are used to detoxify
CC   xenobiotic compounds in plants and seeds. The enzymes are also useful as
CC   targets to facilitate design and/or identify inhibitors of the enzymes
CC   that may be used as herbicides or herbicide synergists. The GST enzymes
CC   produced in the host cells, particularly in microbial host cells, are
CC   useful in preparing antibodies to the enzymes. These antibodies are
CC   useful for detecting the enzymes in situ in cells or in vitro in cell
CC   extracts.
XX
SQ   Sequence   216 AA;


alignment_scores:
            Quality:   39.00              Length:      7
              Ratio:  5.571                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 71.429

alignment_block:
 US-09-696-169-1/rev x AAB03736   ..

 Align seg 1/1  to: AAB03736  from: 1  to: 216

      24 TGGCCTTCCCCATATGGAATT 4
         ||||||||||||:::|||::: 
      11 TrpProSerProPheGlyMet 17

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY79517

seq_documentation_block:
ID   AAY79517 standard; Protein; 216 AA.
XX
AC   AAY79517;
XX
DT   01-AUG-2000  (first entry)
XX
DE   Soybean glutathione-S-transferase se6.pk0048.d7.
XX
KW   Soybean; glutathione-S-transferase; GST; xenobiotic;
KW   detoxification; transgenic plant; herbicide tolerance.
XX
```

```
OS   Glycine max.
XX
PN   WO200018936-A1.
XX
PD   06-APR-2000.
XX
PF   30-SEP-1998;  98WO-US20501.
XX
PR   30-SEP-1998;  98WO-US20501.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   McGonigle B,  O'Keefe DP;
XX
DR   WPI; 2000-317517/27.
DR   N-PSDB; AAZ94954.
XX
PT   Nucleic acids encoding soybean glutathione-S-transferase enzymes useful
PT   for conferring herbicide resistance to plants -
XX
PS   Claim 4; Page 50; 76pp; English.
XX
CC   The present sequence is that a soybean class III
CC   glutathione-S-transferase (GST), as deduced from soybean embryo
CC   cDNA clone se6.pk0048.d7 (see AAZ94954).  The invention provides
CC   soybean GST enzymes (see AAY79512-25) involved in the detoxification
CC   of xenobiotic compounds, especially herbicides, in plants and seeds.
CC   Chimeric genes encoding all or a portion of soybean GST, host cells,
CC   and methods of recombinant production of soybean GST enzymes are
CC   provided.  The sequences are useful in the construction of
CC   herbicide-tolerant transgenic plants, in the recombinant production
CC   of GST enzymes, in the development of screening assays to identify
CC   compounds inhibitory to the GST enzymes (useful as herbicides or
CC   herbicide synergists), and in screening assays to identify chemical
CC   substrates of the GSTs.
XX
SQ   Sequence  216 AA;


alignment_scores:
           Quality:   39.00            Length:      7
             Ratio:  5.571              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 71.429

alignment_block:
 US-09-696-169-1/rev x AAY79517   ..

 Align seg 1/1  to: AAY79517  from: 1  to: 216

      24 TGGCCTTCCCCATATGGAATT 4
         ||||||||||||:::|||:::
      11 TrpProSerProPheGlyMet 17

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB66735

seq_documentation_block:
ID   AAB66735 standard; protein; 216 AA.
XX
AC   AAB66735;
XX
DT   09-APR-2001  (first entry)
XX
DE   Soybean type III GST protein #3.
XX
KW   Soybean; glutathione-S-transferase; herbicide; GST.
XX
OS   Glycine max.
XX
PN   US6171839-B1.
XX
PD   09-JAN-2001.
XX
PF   22-APR-1999;   99US-0296715.
XX
PR   05-SEP-1997;   97US-0924747.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   McGonigle B,  O'Keefe DP;
XX
DR   WPI; 2001-136874/14.
XX
PT   Novel soybean glutathione-S-transferase enzymes useful as targets to
PT   facilitate design and/or identification of inhibitors of the enzyme,
PT   that are used as herbicides or herbicide synergists  -
XX
PS   Claim 1; Column 41-44; 37pp; English.
XX
CC   The present invention relates to soybean glutathione-S-transferase
CC   proteins. The novel sequences are useful in the construction of
CC   herbicide tolerant transgenic plants, in the recombinant production
CC   of glutathione-S-transferase (GST) enzymes, in the development of
CC   screening assays to identify compounds inhibitory to the GST enzymes,
CC   and in screening assays to identify chemical substrates of the GSTs.
XX
SQ   Sequence   216 AA;


alignment_scores:
           Quality:   39.00            Length:      7
             Ratio:  5.571              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 71.429

alignment_block:
 US-09-696-169-1/rev x AAB66735   ..

 Align seg 1/1  to: AAB66735  from: 1  to: 216

      24 TGGCCTTCCCCATATGGAATT 4
         ||||||||||||:::|||:::
      11 TrpProSerProPheGlyMet 17
```

OM of: US-09-696-169-1 to: PIR_68:* out_format : pfs

Date: Nov 5, 2001 7:44 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09696169/runat_05112001_064810_9997/app_query.fasta_1.85
-DB=PIR_68 -QFMT=fastan -SUFFIX=rpr -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=US09696169_@CGN1_1_88 -NCPU=6 -ICPU=3
-LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-696-169-1
Query length: 29
Database: PIR_68:*
Database sequences: 219241
Database length: 76174552
Search time (sec): 22.540000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
pir2:D71475	-	43.00	130.18	10.11	289	probable AMP nucleosidase - Chl
pir2:A81738	-	43.00	130.18	10.11	289	AMP nucleosidase-related protei
pir2:T24853	+	43.00	128.14	10.24	371	hypothetical protein T12A7.4 -
pir2:T15504	-	43.00	124.16	10.48	604	hypothetical protein C15B12.5 -
pir2:T43292	-	43.00	123.17	10.54	682	G protein-linked acetylcholine
pir2:B42512	+	42.00	131.38	15.20	165	G6R protein - vaccinia virus (s
pir2:T28507	+	42.00	131.38	15.20	165	hypothetical protein H6R - vari
pir2:S33083	+	42.00	131.38	15.20	165	G6R protein - variola virus
pir2:T37352	+	42.00	131.38	15.20	165	probable 18.9K protein - vaccin
pir2:C72159	+	42.00	131.38	15.20	165	I7R protein - variola minor vir
pir2:B72264	-	42.00	123.90	15.88	412	hypothetical protein TM1346 - T
pir2:E70514	-	42.00	123.86	15.89	414	probable cysS2 - Mycobacterium
pir2:S23581	+	42.00	123.57	15.91	429	lamB protein precursor - Klebsi
pir2:H86551	-	41.00	137.58	22.19	51	hypothetical protein CPj0492 [im
pir2:C72072	-	41.00	137.58	22.19	51	hypothetical protein CP0262 [imp
pir2:A53808	-	41.00	122.81	24.20	311	homeotic protein cdx-2 - mouse
pir1:YYEC	-	41.00	119.60	24.65	461	cysteine--tRNA ligase (EC 6.1.1
pir2:G85552	-	41.00	119.60	24.65	461	cysteine tRNA synthetase [impor
pir2:H69185	+	41.00	119.37	24.69	474	amidophosphoribosyltransferase
pir2:D86602	-	40.00	119.91	37.25	293	AMP nucleosidase [imported] - C
pir2:A72021	-	40.00	119.91	37.25	293	AMP nucleosidase-related protei
pir1:S62590	+	40.00	116.03	38.10	471	peptidyl-prolyl cis-trans isome
pir2:D83814	-	40.00	115.66	38.19	493	Na+/H+ antiporter BH1316 [impor
pir2:H83691	+	40.00	111.71	39.08	800	hypothetical protein BH0336 [im
pir2:T17685	+	39.00	126.93	54.12	82	hypothetical protein a195R - Chl
pir2:G83895	-	39.00	119.52	56.51	203	hypothetical protein BH1967 [im
pir2:G69901	-	39.00	119.32	56.58	208	acyl-carrier protein phosphodie
pir2:D84132	-	39.00	119.32	56.58	208	acyl-carrier protein phosphodie
pir2:T07156	-	39.00	119.02	56.68	216	probable glutathione transferas
pir2:T09781	-	39.00	118.94	56.71	218	glutathione transferase (EC 2.5
pir2:T06239	-	39.00	118.90	56.72	219	probable glutathione transferas
pir2:T21992	-	39.00	116.01	57.69	312	hypothetical protein F39B2.11 -
pir2:T08353	-	39.00	114.27	58.27	386	hypothetical protein H1549 [imp
pir2:C71420	-	39.00	112.88	58.75	458	hypothetical protein - Arabidop
pir2:H82494	-	39.00	112.19	58.99	498	probable NADH dehydrogenase VCA
pir2:C70606	-	39.00	109.28	60.00	711	probable fadE34 protein - Mycob
pir2:I52657	-	39.00	106.69	60.92	977	seizure-related protein SEZ-6 p
pir2:C71639	-	39.00	105.22	61.44	1169	hypothetical protein RP785 - R
pir2:A71163	-	38.50	108.44	74.18	641	hypothetical protein PH0502 - P
pir2:E69124	-	38.00	115.98	87.34	207	cobalamin biosynthesis protein
pir2:B86873	-	38.00	113.31	88.71	287	oxidoreductase yugB [imported]
pir2:T35337	-	38.00	112.44	89.16	319	probable ion channel subunit
pir2:G65086	-	38.00	111.78	89.50	346	hypothetical protein b3001 -
pir2:E85959	-	38.00	111.78	89.50	346	probable reductase Z4354 [im
pir1:PNFMGF	+	38.00	111.59	89.60	354	peptide-N4-(N-acetyl-beta-gl

seq_name: pir2:D71475

seq_documentation_block:
probable AMP nucleosidase - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C;Species: Chlamydia trachomatis
C;Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C;Accession: D71475
R;Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche
Science 282, 754-759, 1998
A;Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
A;Reference number: A71570; MUID:99000809
A;Accession: D71475
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-289 <ARN>
A;Cross-references: GB:AE001347; GB:AE001273; NID:g3329210; PIDN:AAC68346.1; PID:g332
A;Experimental source: serotype D, strain UW-3/Cx
C;Genetics:
A;Gene: amn

alignment_scores:
Quality: 43.00 Length: 8
Ratio: 5.375 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 87.500

alignment_block:
US-09-696-169-1/rev x D71475 ..

Align seg 1/1 to: D71475 from: 1 to: 289

```
     26 CGTGGCCTTCCCCATATGGAATTC 3
        :::|||||||||||||||||||||
    264 LysGlyLeuProHisMetGluPhe 271
```

seq_name: pir2:A81738

seq_documentation_block:
AMP nucleosidase-related protein TC0132 [imported] - Chlamydia muridarum (strain Nigg
C;Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C;Accession: A81738
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A;Reference number: A81500; MUID:20150255
A;Accession: A81738
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-289 <TET>
A;Cross-references: GB:AE002280; GB:AE002160; NID:g7190162; PIDN:AAF39010.1; PID:g719
A;Experimental source: strain Nigg (MoPn)
C;Genetics:
A;Gene: TC0132

alignment_scores:
Quality: 43.00 Length: 8
Ratio: 5.375 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 87.500

alignment_block:
US-09-696-169-1/rev x A81738 ..

Align seg 1/1 to: A81738 from: 1 to: 289

```
     26 CGTGGCCTTCCCCATATGGAATTC 3
```

```
          :::||||||||||||||||||||
      264 LysGlyLeuProHisMetGluPhe 271

seq_name: pir2:T24853

seq_documentation_block:
hypothetical protein T12A7.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T24853
R;Lennard, N.
submitted to the EMBL Data Library, June 1996
A;Reference number: Z19943
A;Accession: T24853
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-371 <WIL>
A;Cross-references: EMBL:Z73911; PIDN:CAA98139.1; GSPDB:GN00022; CESP:T12A7.4
A;Experimental source: clone T12A7
C;Genetics:
A;Gene: CESP:T12A7.4
A;Map position: 4
A;Introns: 105/3; 138/3; 166/2; 218/1; 261/3; 318/1; 343/3


alignment_scores:
            Quality:   43.00            Length:      8
              Ratio:  6.143              Gaps:      0
 Percent Similarity:  87.500   Percent Identity: 75.000

alignment_block:
 US-09-696-169-1 x T24853   ..

 Align seg 1/1  to: T24853  from: 1  to: 371

       1 GGGAATTCCATATGGGGAAGGCCA 24
         ||||||   :::|||||||||||||
     325 GlyAsnMetValTrpGlyArgPro 332

seq_name: pir2:T15504

seq_documentation_block:
hypothetical protein C15B12.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Mar-2000
C;Accession: T15504
R;Nhan, M.
submitted to the EMBL Data Library, March 1995
A;Description: The sequence of C. elegans cosmid C15B12.
A;Reference number: Z18362
A;Accession: T15504
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-604 <NHA>
A;Cross-references: EMBL:U23529; NID:g746592; PID:g746597; PIDN:AAC46580.1; CESP:C15B12.
A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP:C15B12.5
A;Introns: 24/3; 58/3; 89/1; 128/1; 174/2; 261/3; 292/3; 317/2; 387/3; 424/3; 470/2; 527
C;Superfamily: vertebrate rhodopsin


alignment_scores:
            Quality:   43.00            Length:      9
              Ratio:  5.375              Gaps:      0
 Percent Similarity:  88.889   Percent Identity: 77.778

alignment_block:
 US-09-696-169-1/rev x T15504   ..

 Align seg 1/1  to: T15504  from: 1  to: 604

      27 TCGTGGCCTTCCCCATATGGAATTCCC 1
```

```
            ||||||   |||||||||:::||||||
      14 SerTrpAspSerProTyrSerIlePro 22

seq_name: pir2:T43292

seq_documentation_block:
G protein-linked acetylcholine receptor - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Sep-2000
C;Accession: T43292
R;Lee, Y.S.; Park, Y.S.; Chang, D.J.; Hwang, J.M.; Min, C.K.; Kaang, B.K.; Cho, N.J.
J. Neurochem. 72, 58-65, 1999
A;Title: Cloning and expression of a G protein-linked acetylcholine receptor from Cae
A;Reference number: Z22398; MUID:99101145
A;Accession: T43292
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-682 <LEE>
A;Cross-references: EMBL:AF075245; NID:g4249641; PIDN:AAD13747.1; PID:g4249642
C;Genetics:
A;Gene: gar-1
A;Map position: X
C;Superfamily: vertebrate rhodopsin


alignment_scores:
            Quality:   43.00            Length:      9
              Ratio:  5.375              Gaps:      0
 Percent Similarity:  88.889   Percent Identity: 77.778

alignment_block:
 US-09-696-169-1/rev x T43292   ..

 Align seg 1/1  to: T43292  from: 1  to: 682

      27 TCGTGGCCTTCCCCATATGGAATTCCC 1
         ||||||   |||||||||:::||||||
      14 SerTrpAspSerProTyrSerIlePro 22

seq_name: pir2:B42512

seq_documentation_block:
G6R protein - vaccinia virus (strain Copenhagen)
C;Species: vaccinia virus
A;Note: host Homo sapiens (man)
C;Date: 09-Nov-1990 #sequence_revision 09-Nov-1990 #text_change 20-Jun-2000
C;Accession: B42512
R;Johnson, G.P.
submitted to GenBank, June 1990
A;Reference number: A33172
A;Accession: B42512
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-165 <JOH>
C;Superfamily: vaccinia virus probable 18.9K protein


alignment_scores:
            Quality:   42.00            Length:      9
              Ratio:  5.250              Gaps:      0
 Percent Similarity:  88.889   Percent Identity: 77.778

alignment_block:
 US-09-696-169-1 x B42512   ..

 Align seg 1/1  to: B42512  from: 1  to: 165

       2 GGAATTCCATATGGGGAAGGCCACGAC 28
         ||||||||||||||||   ||||||:::
      97 GlyIleProTyrGlyPheGlyHisAsn 105

seq_name: pir2:T28507
```

```
seq_documentation_block:
hypothetical protein H6R - variola major virus
C;Species: variola major virus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
C;Accession: T28507
R;Massung, R.F.; Esposito, J.J.; Liu, L.I.; Qi, J.; Utterback, T.R.; Knight, J.C.; Aubin
Nature 366, 748-751, 1993
A;Title: Potential virulence determinants in terminal regions of variola smallpox virus
A;Reference number: Z20488; MUID:94088747
A;Accession: T28507
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-165 <MAS>
A;Cross-references: EMBL:L22579; NID:g623595; PIDN:AAA60817.1; PID:g438987
A;Experimental source: strain Bangladesh-1975
C;Superfamily: vaccinia virus probable 18.9K protein


alignment_scores:
            Quality:   42.00             Length:      9
              Ratio:  5.250               Gaps:      0
 Percent Similarity:  88.889   Percent Identity: 77.778

alignment_block:
 US-09-696-169-1 x T28507   ..

 Align seg 1/1  to: T28507  from: 1  to: 165

       2 GGAATTCCATATGGGGAAGGCCACGAC 28
         ||||||||||||||||   ||||||:::
      97 GlyIleProTyrGlyPheGlyHisAsn 105

seq_name: pir2:S33083

seq_documentation_block:
G6R protein - variola virus
N;Alternate names: H6R protein
C;Species: variola virus
C;Date: 02-Dec-1993 #sequence_revision 26-May-1995 #text_change 23-Mar-2001
C;Accession: S33083; C36844
R;Shchelkunov, S.N.; Blinov, V.M.; Totmenin, A.V.; Marennikova, S.S.; Kolykhalov, A.A.;
dzhaparidze, O.G.; Sandakhchiev, L.S.
Virus Res. 27, 25-35, 1993
A;Title: Nucleotide sequence analysis of variola virus HindIII M, L, I genome fragments.
A;Reference number: S33069; MUID:93190624
A;Accession: S33083
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-165 <SHC>
A;Cross-references: EMBL:X67119; NID:g62330; PIDN:CAA47568.1; PID:g62345
A;Experimental source: strain India-1967, isolate Ind3
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1992
R;Blinov, V.M.
submitted to GenBank, November 1992
A;Reference number: A36859
A;Accession: C36844
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-165 <BLI>
A;Cross-references: GB:X69198; NID:g456758; PIDN:CAA49010.1; PID:g297249
A;Experimental source: strain India-1967, ssp. major, isolate Ind3
C;Superfamily: vaccinia virus probable 18.9K protein


alignment_scores:
            Quality:   42.00             Length:      9
              Ratio:  5.250               Gaps:      0
 Percent Similarity:  88.889   Percent Identity: 77.778

alignment_block:
 US-09-696-169-1 x S33083   ..

 Align seg 1/1  to: S33083  from: 1  to: 165

       2 GGAATTCCATATGGGGAAGGCCACGAC 28
         ||||||||||||||||   ||||||:::
      97 GlyIleProTyrGlyPheGlyHisAsn 105

seq_name: pir2:T37352

seq_documentation_block:
probable 18.9K protein - vaccinia virus (strain Ankara)
C;Species: vaccinia virus
A;Variety: strain Ankara
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jun-2000
C;Accession: T37352
R;Antoine, G.; Scheiflinger, F.; Falkner, F.G.; Dorner, F.
submitted to the EMBL Data Library, March 1997
A;Description: The complete genomic sequence of the Modified Vaccinia Ankara (MVA) st
A;Reference number: Z20877
A;Accession: T37352
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-165 <ANT>
A;Cross-references: EMBL:U94848; PIDN:AAB96495.1
A;Experimental source: strain Ankara
C;Genetics:
A;Note: MVA076R
C;Superfamily: vaccinia virus probable 18.9K protein


alignment_scores:
            Quality:   42.00             Length:      9
              Ratio:  5.250               Gaps:      0
 Percent Similarity:  88.889   Percent Identity: 77.778

alignment_block:
 US-09-696-169-1 x T37352   ..

 Align seg 1/1  to: T37352  from: 1  to: 165

       2 GGAATTCCATATGGGGAAGGCCACGAC 28
         ||||||||||||||||   ||||||:::
      97 GlyIleProTyrGlyPheGlyHisAsn 105

seq_name: pir2:C72159

seq_documentation_block:
I7R protein - variola minor virus (strain Garcia-1966)
C;Species: variola minor virus
C;Date: 24-Nov-1999 #sequence_revision 24-Nov-1999 #text_change 20-Jun-2000
C;Accession: C72159
R;Shchelkunov, S.N.; Totmenin, A.V.; Gutorov, V.V.; Safronov, P.F.; Massung, R.F.; Lo
submitted to GenBank, March 1998
A;Description: Analysis of the complete coding sequence of DNA of alastrim variola mi
A;Reference number: A72150
A;Accession: C72159
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-165 <SHC>
A;Cross-references: GB:Y16780; NID:g5830555; PIDN:CAB54669.1; PID:g5830630
A;Experimental source: strain Garcia-1966
C;Genetics:
A;Gene: I7R
C;Superfamily: vaccinia virus probable 18.9K protein


alignment_scores:
            Quality:   42.00             Length:      9
              Ratio:  5.250               Gaps:      0
 Percent Similarity:  88.889   Percent Identity: 77.778

alignment_block:
 US-09-696-169-1 x C72159   ..

 Align seg 1/1  to: C72159  from: 1  to: 165
```

```
      2 GGAATTCCATATGGGGAAGGCCACGAC 28
        ||||||||||||||   ||||||:::
     97 GlyIleProTyrGlyPheGlyHisAsn 105

seq_name: pir2:B72264

seq_documentation_block:
hypothetical protein TM1346 - Thermotoga maritima (strain MSB8)
C;Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C;Accession: B72264
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
 C.M.
Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A;Reference number: A72200; MUID:99287316
A;Accession: B72264
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-412 <ARN>
A;Cross-references: GB:AE001789; GB:AE000512; NID:g4981904; PIDN:AAD36417.1; PID:g498190
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM1346
C;Superfamily: mitochondrial processing peptidase alpha chain


alignment_scores:
            Quality:   42.00             Length:      8
              Ratio:  6.000               Gaps:      0
 Percent Similarity:  87.500   Percent Identity: 75.000

alignment_block:
 US-09-696-169-1/rev x B72264   ..

 Align seg 1/1  to: B72264  from: 1  to: 412

     24 TGGCCTTCCCCATATGGAATTCCC 1
        ||||||:::|||||||||   |||
    145 TrpProGlyProTyrGlyArgPro 152

seq_name: pir2:E70514

seq_documentation_block:
probable cysS2 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: E70514
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987
A;Accession: E70514
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-414 <COL>
A;Cross-references: GB:Z97559; GB:AL123456; NID:g3261820; PIDN:CAB10724.1; PID:g3261823
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: cysS2
C;Superfamily: cysteine--tRNA ligase


alignment_scores:
            Quality:   42.00             Length:      7
              Ratio:  6.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 85.714

alignment_block:
 US-09-696-169-1/rev x E70514   ..

 Align seg 1/1  to: E70514  from: 1  to: 414

     27 TCGTGGCCTTCCCCATATGGA 7
        |||||||||||||||:::|||
    217 SerTrpProSerProPheGly 223

seq_name: pir2:S23581

seq_documentation_block:
lamB protein precursor - Klebsiella pneumoniae
C;Species: Klebsiella pneumoniae
C;Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 20-Aug-1999
C;Accession: S23581
R;Werts, C.; Charbit, A.; Bachellier, S.; Hofnung, M.
Mol. Gen. Genet. 233, 372-378, 1992
A;Title: DNA sequence analysis of the lamB gene from Klebsiella pneumoniae: implicati
A;Reference number: S23581; MUID:92318889
A;Accession: S23581
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-429 <WER>
A;Cross-references: EMBL:X66952; NID:g43816; PIDN:CAA47377.1; PID:g43817
C;Genetics:
A;Gene: lamB
C;Superfamily: lambda receptor protein
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-429/Product: lamB protein #status predicted <MAT>


alignment_scores:
            Quality:   42.00             Length:      8
              Ratio:  5.250               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 75.000

alignment_block:
 US-09-696-169-1 x S23581   ..

 Align seg 1/1  to: S23581  from: 1  to: 429

      1 GGGAATTCCATATGGGGAAGGCCA 24
        |||||||||:::|||:::||||||
    368 GlyAsnSerValTrpSerArgPro 375

seq_name: pir2:H86551

seq_documentation_block:
hypothetical protein CPj0492 [imported] - Chlamydophila pneumoniae (strain J138)
C;Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C;Accession: H86551
R;Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;
Nucleic Acids Res. 28, 2311-2314, 2000
A;Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A;Reference number: A86491; MUID:20330349
A;Accession: H86551
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-51 <STO>
A;Cross-references: GB:BA000008; NID:g8978862; PIDN:BAA98698.1; GSPDB:GN00142
A;Experimental source: strain J138
C;Genetics:
A;Gene: CPj0492


alignment_scores:
            Quality:   41.00             Length:      8
              Ratio:  5.857               Gaps:      0
 Percent Similarity:  87.500   Percent Identity: 75.000

alignment_block:
```

```
 US-09-696-169-1/rev x H86551   ..

 Align seg 1/1  to: H86551  from: 1  to: 51

      24 TGGCCTTCCCCATATGGAATTCCC 1
         |||   ||||||||||||:::|||
      35 TrpSerSerProTyrGlyPhePro 42

seq_name: pir2:C72072

seq_documentation_block:
hypothetical protein CP0262 [imported] - Chlamydophila pneumoniae (strains CWL029 and AR
C;Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C;Accession: C72072; F81596
R;Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A;Title: Comparative genomes of Clamydia pneumoniae and C. trachomatis.
A;Reference number: A72000; MUID:99206606
A;Accession: C72072
A;Molecule type: DNA
A;Residues: 1-51 <ARN>
A;Cross-references: GB:AE001634; GB:AE001363; NID:g4376771; PIDN:AAD18632.1; PID:g437677
A;Experimental source: strain CWL029
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A;Reference number: A81500; MUID:20150255
A;Accession: F81596
A;Molecule type: DNA
A;Residues: 1-51 <REA>
A;Cross-references: GB:AE002186; GB:AE002161; NID:g7189181; PIDN:AAF38124.1; PID:g718918
A;Experimental source: strain AR39, HL cells
C;Genetics:
A;Gene: CPn0492; CP0262


alignment_scores:
           Quality:   41.00              Length:       8
             Ratio:  5.857                Gaps:       0
 Percent Similarity:  87.500   Percent Identity: 75.000

alignment_block:
 US-09-696-169-1/rev x C72072   ..

 Align seg 1/1  to: C72072  from: 1  to: 51

      24 TGGCCTTCCCCATATGGAATTCCC 1
         |||   ||||||||||||:::|||
      35 TrpSerSerProTyrGlyPhePro 42
```

```
OM of: US-09-696-169-1 to: SwissProt_39:*   out_format : pfs

Date: Nov 5, 2001  7:45 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_n2p.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09696169/runat_05112001_064811_10094/app_query.fasta_1.85
 -DB=SwissProt_39 -QFMT=fastan -SUFFIX=rsp -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09696169_@CGN1_1_39 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-696-169-1
Query length: 29
Database: SwissProt_39:*
Database sequences: 93435
Database length: 34255486
Search time (sec): 15.880000


score_list:
Sequence                    Strd Orig    ZScore   EScore Len  ! Documentation  ..
SwissProt_39:YTJ5_CAEEL -    43.00  118.96  9.19      604  ! Q18007 caenorhabditis elegan
SwissProt_39:VG06_VACCC +    42.00  125.98  13.67     165  ! P21027 vaccinia virus (strai
SwissProt_39:VG06_VARV +     42.00  125.98  13.67     165  ! P32996 variola virus. protein
SwissProt_39:RS4_ZYMMO +     42.00  124.32  13.68     204  ! Q9z5u7 zymomonas mobilis. 30s
SwissProt_39:SYC2_MYCTU -    42.00  118.78  13.71     414  ! O33264 mycobacterium tuberc
SwissProt_39:LAMB_KLEPN +    42.00  118.50  13.71     429  ! P31242 klebsiella pneumoniae
SwissProt_39:ACPS_MYCLE +    41.00  124.71  20.41     130  ! Q9x7e3 mycobacterium leprae.
SwissProt_39:CDX2_HUMAN -    41.00  117.88  20.47     311  ! Q99626 homo sapiens (human).
SwissProt_39:CDX2_MOUSE -    41.00  117.88  20.47     311  ! P43241 mus musculus (mouse).
SwissProt_39:SYC_ECOLI -     41.00  114.80  20.50     461  ! P21888 escherichia coli. cyst
SwissProt_39:PUR1_METTH +    41.00  114.59  20.50     474  ! O26742 methanobacterium ther
SwissProt_39:YAL5_SCHPO +    40.00  111.50  30.65     471  ! Q09928 schizosaccharomyces p
SwissProt_39:CU22_BOMMO -    39.00  116.16  45.64     174  ! O02388 bombyx mori (silk mot
SwissProt_39:ACPD_BACSU -    39.00  114.76  45.67     208  ! O35022 bacillus subtilis. pu
SwissProt_39:MTX1_CAEEL -    39.00  111.59  45.74     312  ! O45503 caenorhabditis elegan
SwissProt_39:Y785_RICPR -    39.00  101.26  45.95     1169  ! O05975 rickettsia prowazeki
SwissProt_39:PNGF_FLAME +    38.00  107.47  68.39     354  ! P21163 flavobacterium mening
SwissProt_39:LAMB_YEREN +    38.00  106.51  68.42     400  ! Q56850 yersinia enterocoliti
SwissProt_39:LAMB_ECOLI +    38.00  105.66  68.44     446  ! P02943 escherichia coli. mal
SwissProt_39:LAMB_SALTY +    38.00  105.56  68.45     452  ! P26466 salmonella typhimuriu
SwissProt_39:SYC1_MYCTU -    38.00  105.27  68.46     469  ! P96862 mycobacterium tuberc
SwissProt_39:PTPX_MACNE -    38.00   99.24  68.64     1013  ! O02695 macaca nemestrina (p
SwissProt_39:PTPJ_HUMAN +    38.00   97.07  68.71     1337  ! Q12913 homo sapiens (human)
SwissProt_39:BFR1_SCHPO +    38.00   96.02  68.74     1530  ! P41820 schizosaccharomyces
SwissProt_39:YXEL_BACSU -    37.00  110.31  101.94    165  ! P54951 bacillus subtilis. hy
SwissProt_39:MPT5_MYCLE +    37.00  108.06  102.04    220  ! Q05868 mycobacterium leprae.
SwissProt_39:MNTB_SYNY3 -    37.00  105.48  102.16    306  ! Q55282 synechocystis sp. (st
SwissProt_39:CEL1_AGABI -    37.00  105.13  102.18    320  ! Q00023 agaricus bisporus (cc
SwissProt_39:PRTZ_HORVU +    37.00  103.40  102.26    399  ! P06293 hordeum vulgare (barl
SwissProt_39:Y872_HAEIN +    37.00  102.10  102.32    471  ! Q57491 haemophilus influenza
SwissProt_39:MALQ_AQUAE -    37.00  101.87  102.33    485  ! O66937 aquifex aeolicus. 4-a
SwissProt_39:PUR1_MYCTU +    37.00  101.22  102.36    527  ! O06626 mycobacterium tuberc
SwissProt_39:Y4OA_RHISN +    37.00  100.30  102.40    593  ! P55586 rhizobium sp. (strain
SwissProt_39:INLB_LISMO +    37.00   99.83  102.42    630  ! P25147 listeria monocytogene
SwissProt_39:AFSK_STRCO +    37.00   97.97  102.51    799  ! P54741 streptomyces coelicol
SwissProt_39:CCAA_MOUSE -    37.00   90.17  102.87    2164  ! P97445 mus musculus (mouse)
SwissProt_39:CYPA_CAEEL +    36.00  108.08  152.30    147  ! P52017 caenorhabditis elegan
SwissProt_39:YCAK_ECOLI +    36.00  105.83  152.45    196  ! P43340 escherichia coli. put
SwissProt_39:PSA1_ORYSA -    36.00  103.32  152.63    270  ! P52428 oryza sativa (rice).
SwissProt_39:PS12_ARATH -    36.00  103.12  152.64    277  ! O23712 arabidopsis thaliana
SwissProt_39:PS11_ARATH -    36.00  103.09  152.64    278  ! P34066 arabidopsis thalia
SwissProt_39:TF2D_SPOFR -    36.00  102.32  152.70    307  ! P53361 spodoptera frugipe
SwissProt_39:SSR5_MOUSE -    36.00  101.03  152.78    362  ! O08858 mus musculus (mous
SwissProt_39:SSR5_RAT -      36.00  101.01  152.79    363  ! P30938 rattus norvegicus (r
SwissProt_39:SSR5_HUMAN -    36.00  100.98  152.79    364  ! P35346 homo sapiens (huma

seq_name: SwissProt_39:YTJ5_CAEEL

seq_documentation_block:
ID   YTJ5_CAEEL     STANDARD;      PRT;   604 AA.
AC   Q18007;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   PROBABLE G PROTEIN-COUPLED RECEPTOR C15B12.5.
GN   C15B12.5.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Nhan M.;
RL   Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: NOT KNOWN. PUTATIVE RECEPTOR.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC       MOST SIMILAR TO MUSCARINIC ACETYLCHOLINE RECEPTORS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U23529; AAC46580.1; -.
DR   WormPep; C15B12.5; CE02487.
DR   InterPro; IPR000276; -.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   Hypothetical protein; G-protein coupled receptor; Transmembrane;
KW   Glycoprotein.
FT   TRANSMEM     21     41       POTENTIAL.
FT   TRANSMEM     67     87       POTENTIAL.
FT   TRANSMEM    102    122       POTENTIAL.
FT   TRANSMEM    145    165       POTENTIAL.
FT   TRANSMEM    190    210       POTENTIAL.
FT   TRANSMEM    529    549       POTENTIAL.
FT   CARBOHYD      3      3       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   604 AA;  68199 MW;  3F2899D0E08BAD62 CRC64;


alignment_scores:
            Quality:   43.00              Length:      9
              Ratio:  5.375                Gaps:       0
 Percent Similarity:  88.889   Percent Identity: 77.778

alignment_block:
 US-09-696-169-1/rev x YTJ5_CAEEL   ..

 Align seg 1/1  to: YTJ5_CAEEL  from: 1  to: 604

      27 TCGTGGCCTTCCCCATATGGAATTCCC 1
         ||||||   |||||||||:::||||||
      14 SerTrpAspSerProTyrSerIlePro 22

seq_name: SwissProt_39:VG06_VACCC

seq_documentation_block:
```

```
ID   VG06_VACCC     STANDARD;      PRT;   165 AA.
AC   P21027;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   PROTEIN G6.
GN   G6R.
OS   Vaccinia virus (strain Copenhagen).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10249;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91021027; PubMed=2219722;
RA   Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA   Paoletti E.;
RT   "The complete DNA sequence of vaccinia virus.";
RL   Virology 179:247-266(1990).
RN   [2]
RP   COMPLETE GENOME.
RA   Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA   Paoletti E.;
RL   Virology 179:517-563(1990).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M35027; AAA48070.1; -.
DR   PIR; B42512; B42512.
SQ   SEQUENCE   165 AA;  18949 MW;  EFE3210AE28E17FE CRC64;


alignment_scores:
            Quality:   42.00              Length:      9
              Ratio:  5.250                Gaps:       0
 Percent Similarity:  88.889   Percent Identity: 77.778

alignment_block:
 US-09-696-169-1 x VG06_VACCC   ..

 Align seg 1/1  to: VG06_VACCC  from: 1  to: 165

       2 GGAATTCCATATGGGGAAGGCCACGAC 28
         ||||||||||||||   ||||||:::
      97 GlyIleProTyrGlyPheGlyHisAsn 105

seq_name: SwissProt_39:VG06_VARV

seq_documentation_block:
ID   VG06_VARV      STANDARD;      PRT;   165 AA.
AC   P32996;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   PROTEIN G6.
GN   G6R OR H6R.
OS   Variola virus.
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10255;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=INDIA-1967 / ISOLATE IND3;
RX   MEDLINE=94152154; PubMed=8109158;
RA   Shchelkunov S.N., Blinov V.M., Resenchuk S.M., Totmenin A.V.,
RA   Sandakhchiev L.S.;
RT   "Analysis of the nucleotide sequence of a 43 kbp segment of the
RT   genome of variola virus India-1967 strain.";
RL   Virus Res. 30:239-258(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=INDIA-1967 / ISOLATE IND3;
RX   MEDLINE=93190624; PubMed=8383392;
RA   Shchelkunov S.N., Blinov V.M., Totmenin A.V., Marennikova S.S.,
RA   Kolykhalov A.A., Frolov I.V., Chizhikov V.E., Gytorov V.V.,
RA   Gashikov P.V., Belanov E.F., Belavin P.A., Resenchuk S.M.,
RA   Andzhaparidze O.G., Sandakhchiev L.S.;
RT   "Nucleotide sequence analysis of variola virus HindIII M, L, I genome
RT   fragments.";
RL   Virus Res. 27:25-35(1993).
RN   [3]
RP   COMPLETE GENOME.
RC   STRAIN=INDIA-1967 / ISOLATE IND3;
RX   MEDLINE=93202281; PubMed=8384129;
RA   Shchelkunov S.N., Blinov V.M., Sandakhchiev L.S.;
RT   "Genes of variola and vaccinia viruses necessary to overcome the host
RT   protective mechanisms.";
RL   FEBS Lett. 319:80-83(1993).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BANGLADESH-1975;
RX   MEDLINE=94088747; PubMed=8264798;
RA   Massung R.F., Esposito J.J., Liu L., Qi J., Utterback T.R.,
RA   Knight J.C., Aubin L., Yuran T.E., Parsons J.M., Loparev V.N.,
RA   Selivanov N.A., Cavallaro K.F., Kerlavage A.R., Mahy B.W.J.,
RA   Venter C.J.;
RT   "Potential virulence determinants in terminal regions of variola
RT   smallpox virus genome.";
RL   Nature 366:748-751(1993).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X67119; CAA47568.1; -.
DR   EMBL; X69198; CAA49010.1; -.
DR   EMBL; L22579; AAA60817.1; -.
DR   PIR; C36844; C36844.
DR   PIR; S33083; S33083.
SQ   SEQUENCE   165 AA;  18934 MW;  CB70D9900518C80E CRC64;


alignment_scores:
            Quality:   42.00              Length:      9
              Ratio:  5.250                Gaps:       0
 Percent Similarity:  88.889   Percent Identity: 77.778

alignment_block:
 US-09-696-169-1 x VG06_VARV   ..

 Align seg 1/1  to: VG06_VARV  from: 1  to: 165

       2 GGAATTCCATATGGGGAAGGCCACGAC 28
         ||||||||||||||   ||||||:::
      97 GlyIleProTyrGlyPheGlyHisAsn 105

seq_name: SwissProt_39:RS4_ZYMMO

seq_documentation_block:
ID   RS4_ZYMMO      STANDARD;      PRT;   204 AA.
AC   Q9Z5U7;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   30S RIBOSOMAL PROTEIN S4.
GN   RPSD.
OS   Zymomonas mobilis.
```

```
CC   -!- INDUCTION: BY MALTOSE.
CC   -!- SIMILARITY: BELONGS TO THE LAMB FAMILY OF PORINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X66952; CAA47377.1; -.
DR   PIR; S23581; S23581.
DR   HSSP; P26466; 1MPR.
KW   Sugar transport; Outer membrane; Transmembrane; Porin; Signal.
FT   SIGNAL        1     25       BY SIMILARITY.
FT   CHAIN        26    429       MALTOPORIN.
FT   DISULFID     47     63       BY SIMILARITY.
FT   SITE        413    415       CELL ATTACHMENT SITE (POTENTIAL).
SQ   SEQUENCE   429 AA;  47804 MW;  C5A51C034A7193B2 CRC64;


alignment_scores:
            Quality:   42.00              Length:      8
              Ratio:  5.250                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 75.000

alignment_block:
 US-09-696-169-1 x LAMB_KLEPN   ..

 Align seg 1/1  to: LAMB_KLEPN  from: 1  to: 429

       1 GGGAATTCCATATGGGGAAGGCCA 24
         ||||||||:::|||:::||||||
     368 GlyAsnSerValTrpSerArgPro 375

seq_name: SwissProt_39:ACPS_MYCLE

seq_documentation_block:
ID   ACPS_MYCLE     STANDARD;      PRT;   130 AA.
AC   Q9X7E3;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   HOLO-[ACYL-CARRIER PROTEIN] SYNTHASE (EC 2.7.8.7) (HOLO-ACP SYNTHASE).
GN   ACPS OR MLCB458.07.
OS   Mycobacterium leprae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Seeger K.J., Harris D., James K.D., Parkhill J., Barrell B.G.,
RA   Rajandream M.A.;
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: TRANSFERS THE 4'-PHOSPHOPANTETHEINE MOIETY FROM COENZYME
CC       A TO A SER OF ACYL-CARRIER PROTEIN (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: COA + APO-[ACYL-CARRIER PROTEIN] = ADENOSINE
CC       3',5'-BISPHOSPHATE + HOLO-[ACYL-CARRIER PROTEIN].
CC   -!- SIMILARITY: BELONGS TO THE ACPS FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL049478; CAB39572.1; -.
DR   InterPro; IPR002582; -.
DR   Pfam; PF01648; ACPS; 1.
KW   Lipid synthesis; Transferase.

SQ   SEQUENCE   130 AA;  14137 MW;  6C747E38BA57D03B CRC64;


alignment_scores:
            Quality:   41.00              Length:      8
              Ratio:  5.125                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 75.000

alignment_block:
 US-09-696-169-1 x ACPS_MYCLE   ..

 Align seg 1/1  to: ACPS_MYCLE  from: 1  to: 130

       4 AATTCCATATGGGGAAGGCCACGA 27
         |||::::::|||||||||||||||
      87 AsnAspMetTrpGlyArgProArg 94

seq_name: SwissProt_39:CDX2_HUMAN

seq_documentation_block:
ID   CDX2_HUMAN     STANDARD;      PRT;   311 AA.
AC   Q99626; O00503;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   HOMEOBOX PROTEIN CDX-2 (CAUDAL-TYPE HOMEOBOX PROTEIN 2) (CDX-3).
GN   CDX2 OR CDX3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon carcinoma;
RX   MEDLINE=97188282; PubMed=9036867;
RA   Mallo G.V., Rechreche H., Frigerio J.M., Rocha D., Zweibaum A.,
RA   Lacasa M., Jordan B.R., Dusetti N.J., Dagorn J.C., Iovanna J.L.;
RT   "Molecular cloning, sequencing and expression of the mRNA encoding
RT   human Cdx1 and Cdx2 homeobox. Down-regulation of Cdx1 and Cdx2 mRNA
RT   expression during colorectal carcinogenesis.";
RL   Int. J. Cancer 74:35-44(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98120622; PubMed=9459001;
RA   Drummond F.J., Putt W., Fox M., Edwards Y.H.;
RT   "Cloning and chromosome assignment of the human CDX2 gene.";
RL   Ann. Hum. Genet. 61:393-400(1997).
CC   -!- FUNCTION: MAY BE NECESSARY FOR SOME GENERAL ASPECT OF COLONIC
CC       EPITHELIAL PHENOTYPE (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: BELONGS TO THE CAUDAL FAMILY OF HOMEOBOX PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U51096; AAB40603.1; -.
DR   EMBL; Y13709; CAA74038.1; -.
DR   HSSP; P02833; 1SAN.
DR   MIM; 600297; -.
DR   InterPro; IPR000047; -.
DR   InterPro; IPR001356; -.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00024; HOMEOBOX.
DR   PRINTS; PR00031; HTHREPRESSR.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT   DOMAIN       46     51       POLY-ALA.
```

```
OC   Bacteria; Proteobacteria; alpha subdivision; Sphingomonas group;
OC   Zymomonas.
OX   NCBI_TaxID=542;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 31821 / ZM4 / CP4;
RA   Um H.W., Kang H.S.;
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 16S RIBOSOMAL RNA
CC       (BY SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 1 S4 RNA-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE S4P FAMILY OF RIBOSOMAL PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF124349; AAD19714.1; -.
DR   InterPro; IPR001912; -.
DR   InterPro; IPR002942; -.
DR   Pfam; PF00163; Ribosomal_S4; 1.
DR   Pfam; PF01479; S4; 1.
DR   PROSITE; PS00632; RIBOSOMAL_S4; 1.
KW   Ribosomal protein; rRNA-binding.
FT   DOMAIN       94    141       RNA-BINDING (S4 TYPE).
SQ   SEQUENCE   204 AA;  23367 MW;  6D4527E0A5E45838 CRC64;


alignment_scores:
            Quality:   42.00              Length:      9
              Ratio:  4.667                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 66.667

alignment_block:
 US-09-696-169-1 x RS4_ZYMMO   ..

 Align seg 1/1  to: RS4_ZYMMO  from: 1  to: 204

       1 GGGAATTCCATATGGGGAAGGCCACGA 27
         |||:::::|||||||||||||||:::
      16 GlyGluAsnIleTrpGlyArgProLys 24

seq_name: SwissProt_39:SYC2_MYCTU

seq_documentation_block:
ID   SYC2_MYCTU     STANDARD;      PRT;   414 AA.
AC   O33264;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CYSTEINYL-TRNA SYNTHETASE 2 (EC 6.1.1.16) (CYSTEINE--TRNA LIGASE 2)
DE   (CYSRS 2).
GN   CYSS2 OR RV2130C OR MTCY261.29C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Sqares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- CATALYTIC ACTIVITY: ATP + L-CYSTEINE + TRNA(CYS) = AMP +
CC       PYROPHOSPHATE + L-CYSTEINYL-TRNA(CYS).
CC   -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC       STRONG, TO METHIONYL-TRNA SYNTHETASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z97559; CAB10724.1; -.
DR   TubercuList; Rv2130c; -.
DR   InterPro; IPR001412; -.
DR   InterPro; IPR002308; -.
DR   Pfam; PF01406; tRNA-synt_1e; 1.
DR   PRINTS; PR00983; TRNASYNTHCYS.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; FALSE_NEG.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
FT   SITE         46     56       "HIGH" REGION.
FT   SITE        291    295       "KMSKS" REGION.
FT   BINDING     294    294       ATP (BY SIMILARITY).
SQ   SEQUENCE   414 AA;  45594 MW;  B03159DB99B871E7 CRC64;


alignment_scores:
            Quality:   42.00              Length:      7
              Ratio:  6.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 85.714

alignment_block:
 US-09-696-169-1/rev x SYC2_MYCTU   ..

 Align seg 1/1  to: SYC2_MYCTU  from: 1  to: 414

      27 TCGTGGCCTTCCCCATATGGA 7
         |||||||||||||||:::|||
     217 SerTrpProSerProPheGly 223

seq_name: SwissProt_39:LAMB_KLEPN

seq_documentation_block:
ID   LAMB_KLEPN     STANDARD;      PRT;   429 AA.
AC   P31242;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   MALTOPORIN PRECURSOR (MALTOSE-INDUCIBLE PORIN).
GN   LAMB.
OS   Klebsiella pneumoniae.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Klebsiella.
OX   NCBI_TaxID=573;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1033-5P14 / KAY2026;
RX   MEDLINE=92318889; PubMed=1535683;
RA   Werts C., Charbit A., Bachellier S., Hofnung M.;
RT   "DNA sequence analysis of the lamB gene from Klebsiella pneumoniae:
RT   implications for the topology and the pore functions in maltoporin.";
RL   Mol. Gen. Genet. 233:372-378(1992).
CC   -!- FUNCTION: INVOLVED IN THE TRANSPORT OF MALTOSE AND MALTODEXTRINS.
CC       DOES NOT ACT AS A RECEPTOR FOR PHAGES.
CC   -!- SUBUNIT: HOMOTRIMER.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC       SPANS THE MEMBRANE 18 TIMES (BY SIMILARITY).
```

```
FT   DOMAIN       84     91       POLY-ALA.
FT   DNA_BIND    184    243       HOMEOBOX.
FT   DOMAIN      248    254       POLY-GLN.
FT   DOMAIN      255    268       POLY-PRO.
FT   CONFLICT     52     52       Q -> AA (IN REF. 2).
FT   CONFLICT     87     87       A -> AA (IN REF. 2).
FT   CONFLICT     93     93       A -> G (IN REF. 2).
SQ   SEQUENCE   311 AA;  33438 MW;  C2FEDEF1089D2367 CRC64;


alignment_scores:
            Quality:   41.00             Length:      9
              Ratio:  5.857               Gaps:       0
 Percent Similarity:  77.778   Percent Identity: 66.667

alignment_block:
 US-09-696-169-1/rev x CDX2_HUMAN   ..

 Align seg 1/1  to: CDX2_HUMAN  from: 1  to: 311

      27 TCGTGGCCTTCCCCATATGGAATTCCC 1
         |||||||||:::   ||||||   |||
      63 SerTrpProAlaAlaTyrGlyAlaPro 71

seq_name: SwissProt_39:CDX2_MOUSE

seq_documentation_block:
ID   CDX2_MOUSE     STANDARD;      PRT;   311 AA.
AC   P43241;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   HOMEOBOX PROTEIN CDX-2 (CAUDAL-TYPE HOMEOBOX PROTEIN 2).
GN   CDX2 OR CDX-2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C;
RX   MEDLINE=94253086; PubMed=7910823;
RA   James R.J., Erler T., Kazenwadel J.;
RT   "Structure of the murine homeobox gene cdx-2. Expression in embryonic
RT   and adult intestinal epithelium.";
RL   J. Biol. Chem. 269:15229-15237(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Small intestine;
RX   MEDLINE=95021263; PubMed=7935448;
RA   Suh E., Chen L., Taylor J., Traber P.G.;
RT   "A homeodomain protein related to caudal regulates intestine-specific
RT   gene transcription.";
RL   Mol. Cell. Biol. 14:7340-7351(1994).
RN   [3]
RP   SEQUENCE OF 204-229 FROM N.A.
RX   MEDLINE=91131633; PubMed=1671571;
RA   James R.J., Kazenwadel J.;
RT   "Homeobox gene expression in the intestinal epithelium of adult
RT   mice.";
RL   J. Biol. Chem. 266:3246-3251(1991).
CC   -!- FUNCTION: MAY BE NECESSARY FOR SOME GENERAL ASPECT OF COLONIC
CC       EPITHELIAL PHENOTYPE.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- TISSUE SPECIFICITY: INTESTINE; EXPRESSED SPECIFICALLY IN GUT
CC       EPITHELIUM WHERE IT IS NOT RESTRICTED TO A PARTICULAR CELL
CC       LINEAGE. ABUNDANT EXPRESSION IS SEEN IN THE PROXIMAL COLON WITH
CC       SLIGHTLY LOWER LEVELS IN DISTAL COLON. EXPRESSION IN THE PROXIMAL
CC       COLON IS NOT RESTRICTED EITHER TO A PARTICULAR CELL LINEAGE OR
CC       STAGE OF DIFFERENTIATION WHILE IN THE DISTAL COLON IT IS MORE
CC       ABUNDANT IN THE DIFFERENTIATED CELLS TOWARDS THE TOP OF THE CRYPT.
CC   -!- SIMILARITY: BELONGS TO THE CAUDAL FAMILY OF HOMEOBOX PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U00454; AAA19645.1; -.
DR   EMBL; S74520; AAB32251.1; -.
DR   HSSP; P02833; 1SAN.
DR   TRANSFAC; T02002; -.
DR   MGD; MGI:88361; Cdx2.
DR   InterPro; IPR000047; -.
DR   InterPro; IPR001356; -.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00024; HOMEOBOX.
DR   PRINTS; PR00031; HTHREPRESSR.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT   DOMAIN       46     53       POLY-ALA.
FT   DOMAIN       85     92       POLY-ALA.
FT   DNA_BIND    185    244       HOMEOBOX.
FT   DOMAIN      247    257       POLY-GLN.
FT   CONFLICT     69     69       Y -> H (IN REF. 2).
SQ   SEQUENCE   311 AA;  33476 MW;  71FFC4C263462FF3 CRC64;


alignment_scores:
            Quality:   41.00             Length:      9
              Ratio:  5.857               Gaps:       0
 Percent Similarity:  77.778   Percent Identity: 66.667

alignment_block:
 US-09-696-169-1/rev x CDX2_MOUSE   ..

 Align seg 1/1  to: CDX2_MOUSE  from: 1  to: 311

      27 TCGTGGCCTTCCCCATATGGAATTCCC 1
         |||||||||:::   ||||||   |||
      64 SerTrpProThrAlaTyrGlyAlaPro 72

seq_name: SwissProt_39:SYC_ECOLI

seq_documentation_block:
ID   SYC_ECOLI      STANDARD;      PRT;   461 AA.
AC   P21888;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   CYSTEINYL-TRNA SYNTHETASE (EC 6.1.1.16) (CYSTEINE--TRNA LIGASE)
DE   (CYSRS).
GN   CYSS.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
RC   STRAIN=K12;
RX   MEDLINE=91195046; PubMed=2014166;
RA   Eriani G., Dirheimer G., Gangloff J.;
RT   "Cysteinyl-tRNA synthetase: determination of the last E. coli
RT   aminoacyl-tRNA synthetase primary structure.";
RL   Nucleic Acids Res. 19:265-269(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91126117; PubMed=1992490;
RA   Hou Y.M., Shiba K., Mottes C., Schimmel P.;
RT   "Sequence determination and modeling of structural motifs for the
RT   smallest monomeric aminoacyl-tRNA synthetase.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:976-980(1991).
```

```
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=91323511; PubMed=1864365;
RA   Avalos J., Corrochano L.M., Brenner S.;
RT   "Cysteinyl-tRNA synthetase is a direct descendant of the first
RT   aminoacyl-tRNA synthetase.";
RL   FEBS Lett. 286:176-180(1991).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,
RA   Duncan M., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA   Lew H., Lin D., Namath A., Oefner P., Schramm S., Davis R.W.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: ATP + L-CYSTEINE + TRNA(CYS) = AMP +
CC       PYROPHOSPHATE + L-CYSTEINYL-TRNA(CYS).
CC   -!- SUBUNIT: MONOMER.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC       STRONG, TO METHIONYL-TRNA SYNTHETASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X56234; CAA39691.1; -.
DR   EMBL; M59381; AAA23658.1; -.
DR   EMBL; X59293; CAA41983.1; -.
DR   EMBL; AE000158; AAC73628.1; -.
DR   EMBL; U82664; AAB40279.1; -.
DR   PIR; A37868; YYEC.
DR   EcoGene; EG10196; cysS.
DR   InterPro; IPR001412; -.
DR   InterPro; IPR002308; -.
DR   Pfam; PF01406; tRNA-synt_1e; 1.
DR   PRINTS; PR00983; TRNASYNTHCYS.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; FALSE_NEG.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
FT   SITE         29     40       "HIGH" REGION.
FT   SITE        266    270       "KMSKS" REGION.
FT   BINDING     269    269       ATP (BY SIMILARITY).
FT   CONFLICT    316    316       L -> V (IN REF. 1).
SQ   SEQUENCE   461 AA;  52202 MW;  2FA77FDBB7C5BA99 CRC64;


alignment_scores:
            Quality:   41.00              Length:      7
              Ratio:  5.857                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 85.714

alignment_block:
 US-09-696-169-1/rev x SYC_ECOLI   ..

 Align seg 1/1  to: SYC_ECOLI  from: 1  to: 461

      27 TCGTGGCCTTCCCCATATGGA 7
         |||||||||||||||:::|||
     193 SerTrpProSerProTrpGly 199
```

```
seq_name: SwissProt_39:PUR1_METTH

seq_documentation_block:
ID   PUR1_METTH     STANDARD;      PRT;   474 AA.
AC   O26742;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   AMIDOPHOSPHORIBOSYLTRANSFERASE PRECURSOR (EC 2.4.2.14) (GLUTAMINE
DE   PHOSPHORIBOSYLPYROPHOSPHATE AMIDOTRANSFERASE) (ATASE) (GPATASE).
GN   PURF OR MTH646.
OS   Methanobacterium thermoautotrophicum.
OC   Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC   Methanothermobacter.
OX   NCBI_TaxID=145262;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DELTA H;
RX   MEDLINE=98037514; PubMed=9371463;
RA   Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA   Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA   Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA   Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA   Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA   McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA   Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RT   "Complete genome sequence of Methanobacterium thermoautotrophicum
RT   deltaH: functional analysis and comparative genomics.";
RL   J. Bacteriol. 179:7135-7155(1997).
CC   -!- CATALYTIC ACTIVITY: 5-PHOSPHO-BETA-D-RIBOSYLAMINE + PYROPHOSPHATE
CC       + L-GLUTAMATE = L-GLUTAMINE + 5-PHOSPHO-ALPHA-D-RIBOSE
CC       1-DIPHOSPHATE + H(2)O (BY SIMILARITY).
CC   -!- PATHWAY: FIRST STEP IN DE NOVO PURINE BIOSYNTHESIS
CC       (BY SIMILARITY).
CC   -!- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
CC       AMIDOTRANSFERASES.
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE
CC       PURINE/PYRIMIDINE PHOSPHORIBOSYLTRANSFERASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000845; AAB85151.1; -.
DR   InterPro; IPR000583; -.
DR   InterPro; IPR000836; -.
DR   InterPro; IPR002375; -.
DR   Pfam; PF00310; GATase_2; 1.
DR   Pfam; PF00156; Pribosyltran; 1.
DR   PROSITE; PS00103; PUR_PYR_PR_TRANSFER; 1.
DR   PROSITE; PS00443; GATASE_TYPE_II; 1.
KW   Purine biosynthesis; Transferase; Glycosyltransferase;
KW   Glutamine amidotransferase.
FT   PROPEP        1     10       BY SIMILARITY.
FT   CHAIN        11    474       AMIDOPHOSPHORIBOSYLTRANSFERASE.
FT   ACT_SITE     11     11       GATASE (BY SIMILARITY).
SQ   SEQUENCE   474 AA;  52660 MW;  16BAF93BBF15A0D2 CRC64;


alignment_scores:
            Quality:   41.00              Length:      7
              Ratio:  5.857                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-696-169-1 x PUR1_METTH   ..

 Align seg 1/1  to: PUR1_METTH  from: 1  to: 474
```

```
      2 GGAATTCCATATGGGGAAGGC 22
        ||||||||||||||||||||
    310 GlyIleProTyrGlyGluGly 316

seq_name: SwissProt_39:YAL5_SCHPO

seq_documentation_block:
ID   YAL5_SCHPO     STANDARD;      PRT;   471 AA.
AC   Q09928;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   PROBABLE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE C21E11.05C (EC 5.2.1.8).
GN   SPAC21E11.05C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   McLean J., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC
CC       PEPTIDE BONDS IN OLIGOPEPTIDES.
CC   -!- SIMILARITY: BELONGS TO THE CYCLOPHILIN-TYPE PPIASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z67999; CAA91964.1; -.
DR   HSSP; Q27450; 1A58.
DR   InterPro; IPR002130; -.
DR   Pfam; PF00160; pro_isomerase; 1.
DR   PRINTS; PR00153; CSAPPISMRASE.
DR   PROSITE; PS00170; CSA_PPIASE_1; 1.
DR   PROSITE; PS50072; CSA_PPIASE_2; 1.
KW   Hypothetical protein; Isomerase; Rotamase.
FT   DOMAIN      237    383       PPIASE, CYCLOPHILIN-TYPE.
SQ   SEQUENCE   471 AA;  53573 MW;  66155D062039F8BB CRC64;


alignment_scores:
            Quality:   40.00             Length:      8
              Ratio:  5.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 75.000

alignment_block:
 US-09-696-169-1 x YAL5_SCHPO   ..

 Align seg 1/1  to: YAL5_SCHPO  from: 1  to: 471

      1 GGGAATTCCATATGGGGAAGGCCA 24
        |||:::|||||||||||||:::|||
    293 GlyGlnSerIleTrpGlyLysPro 300

seq_name: SwissProt_39:CU22_BOMMO

seq_documentation_block:
ID   CU22_BOMMO     STANDARD;      PRT;   174 AA.
AC   O02388;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   LARVAL CUTICLE PROTEIN LCP-22 PRECURSOR.
GN   LCP22.
OS   Bombyx mori (Silk moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Bombycoidea; Bombycidae; Bombyx.
OX   NCBI_TaxID=7091;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KINSHU X SHOWA; TISSUE=Epidermis;
RX   MEDLINE=98105581; PubMed=9443370;
RA   Nakato H., Takekoshi M., Togawa T., Izumi S., Tomino S.;
RT   "Purification and cDNA cloning of evolutionally conserved larval
RT   cuticle proteins of the silkworm, Bombyx mori.";
RL   Insect Biochem. Mol. Biol. 27:701-709(1997).
CC   -!- FUNCTION: COMPONENT OF THE CUTICLE OF THE LARVA OF BOMBYX MORI.
CC   -!- SIMILARITY: CONTAINS A CUTICLE CONSENSUS DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB004767; BAA20475.1; -.
DR   InterPro; IPR000618; -.
DR   Pfam; PF00379; insect_cuticle; 1.
DR   PRINTS; PR00947; CUTICLE.
DR   PROSITE; PS00233; CUTICLE; 1.
KW   Structural protein; Cuticle; Signal.
FT   SIGNAL        1     16       POTENTIAL.
FT   CHAIN        17    174       LARVAL CUTICLE PROTEIN LCP-22.
SQ   SEQUENCE   174 AA;  18852 MW;  A54DBA63F17943F5 CRC64;


alignment_scores:
            Quality:   39.00             Length:      9
              Ratio:  5.571               Gaps:      0
 Percent Similarity:  77.778   Percent Identity: 66.667

alignment_block:
 US-09-696-169-1/rev x CU22_BOMMO   ..

 Align seg 1/1  to: CU22_BOMMO  from: 1  to: 174

     27 TCGTGGCCTTCCCCATATGGAATTCCC 1
        ||||||   ||||||   |||:::|||
    122 SerTrpThrSerProGluGlyValPro 130

seq_name: SwissProt_39:ACPD_BACSU

seq_documentation_block:
ID   ACPD_BACSU     STANDARD;      PRT;   208 AA.
AC   O35022;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   PUTATIVE ACYL CARRIER PROTEIN PHOSPHODIESTERASE (EC 3.1.4.14) (ACP
DE   PHOSPHODIESTERASE).
GN   ACPD.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;
RT   "Sequence analysis of the Bacillus subtilis chromosome region between
RT   the terC and odhAB loci cloned in a yeast artificial chromosome.";
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CONVERTS HOLO-ACP TO APO-ACP BY HYDROLYTIC CLEAVAGE OF
CC       THE PHOSPHOPANTETHEINE RESIDUE FROM ACP (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: [ACYL-CARRIER PROTEIN] + H(2)O = 4'-
CC       PHOSPHOPANTETHEINE + APOPROTEIN.
```

```
CC   -!- SIMILARITY: BELONGS TO THE ACPD FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF027868; AAB84476.1; -.
DR   EMBL; Z99114; CAB13815.1; -.
DR   SubtiList; BG13523; acpD.
KW   Hydrolase.
SQ   SEQUENCE   208 AA;  22977 MW;  946C054CF044336C CRC64;


alignment_scores:
           Quality:   39.00             Length:      8
             Ratio:  5.571               Gaps:      0
 Percent Similarity:  87.500   Percent Identity: 75.000

alignment_block:
 US-09-696-169-1/rev x ACPD_BACSU   ..

 Align seg 1/1  to: ACPD_BACSU  from: 1  to: 208

      28 GTCGTGGCCTTCCCCATATGGAAT 5
         |||   |||||||||:::|||||
      94 ValPheAlaPheProLeuTrpAsn 101

seq_name: SwissProt_39:MTX1_CAEEL

seq_documentation_block:
ID   MTX1_CAEEL     STANDARD;      PRT;   312 AA.
AC   O45503;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   METAXIN 1 HOMOLOG.
GN   F39B2.11.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Dobson R.;
RL   Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INVOLVED IN TRANSPORT OF PROTEINS INTO THE
CC       MITOCHONDRION. ESSENTIAL FOR EMBRYONIC DEVELOPMENT (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL OUTER MEMBRANE (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE METAXIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z92834; CAB07391.1; -.
DR   WormPep; F39B2.11; CE16016.
KW   Hypothetical protein; Mitochondrion; Outer membrane; Transmembrane;
KW   Transport; Protein transport.
FT   TRANSMEM    282    302       POTENTIAL.
SQ   SEQUENCE   312 AA;  35319 MW;  C870D8BCE4CB7EB3 CRC64;


alignment_scores:
           Quality:   39.00             Length:      8
             Ratio:  5.571               Gaps:      0
 Percent Similarity:  87.500   Percent Identity: 62.500

alignment_block:
 US-09-696-169-1/rev x MTX1_CAEEL   ..

 Align seg 1/1  to: MTX1_CAEEL  from: 1  to: 312

      24 TGGCCTTCCCCATATGGAATTCCC 1
         |||||||||   :::|||:::|||
       6 TrpProSerAspPheGlyLeuPro 13
```

STIC-ILL

QR 180.J6

From: Huynh, Phuong N.
Sent: Monday, November 05, 2001 2:46 PM
To: STIC-ILL
Subject: RE: 09/696,169

Please deliver the following:

Int Arch Allergy Immunol 108: 55-?; 1995

J Immunol 151: 4773-?; 1993

J Immunol 163(10): 5489-96; 1999

J Immunol 162(4): 2406-14;

Thanks,

Neon
Art unit 1644
Mail CM1, 9E12
Office CM1, 9D06
Tel 308-4844